International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Detecting Malicious Posts in Social Networks Using Text Analysis

Neeraja M[1], John Prakash[2]

[1]PG Scholar, Dept. of Computer Science and Engineering, MITE Moodabidri, India

[2]Senior Assistant Professor, Dept. of Computer Science and Engineering, MITE Moodabidri, India

Abstract: ***We have reached the era of social media networks represented by Facebook, Twitter, Flicker and YouTube. Internet users spend most of their time on social networks than search engines. Public figures and business entities set up social networking pages to promote direct interactions with the online users. Social media systems heavily depend on users for getting content and sharing. Information used is spread across the social networks in quick and effective manner. However, at the same time social media networks become vulnerable to different types of unwanted and malicious hacker or spammer actions. It has been observed that there is a greater participation in Facebook pages regarding malicious content generation. These contents will be in greater amount as compared to legitimate content. In this work we develop a detection mechanism to distinguish between malicious and genuine posts within seconds after the posts are uploaded by user. This work proposes an extensive keyword set based on the textual content and URL features to identify malicious content on Facebook at zero time. The intent is to catch malicious or vulgar content that is currently evading Facebook's detection mechanisms.***

Keywords: Social Network, Text Analysis, Similarity, Suspicious Post, Suspicious URL, Cybercriminals

1. Introduction

Social network sites like Facebook, Twitter, and Google+ are experiencing incredible growth in users. There are more than a million users as of now. Besides just creating a profile and linking with friends, the social networks are now building platforms to run their website. These platforms are built based on the user profile details. These social applications are soon becoming an example of online communication which makes use of the user's private information and activities in social links for various services. The Social networks are popular means of communication among the internet users.

People are heavily relaying on online interactions. The internet is giving different options to create and maintain contacts and relations for the user. With the introduction of social media network these options have become even easier to be used. Due to this heavy use of social media network a certain group of internet users called cybercriminal make use of this opportunity for threads. Cybercriminals use different means to create spams fraud and other attacks on the users. Another means of attack by cybercriminals is the misuse of videos, images and links showed by the user.

Cyber attacks primarily occur on social networks. Popular sites such as Facebook and Twitter currently have millions of active users. The popularity of social networks makes them exiting venues to for executing malicious activities. Due to the huge popularity of social media network these makes it easy for cybercriminal to misuse them. These can be in the form of media, thread or malicious post which does not belongs to a user. These post upon clicking will take the user to some other pages created by malicious user.

Cybercriminals create interesting posts that are actually baits which will be attracted by some users. Typical social engineering plans include the use of Interesting posts that ride on seasonal events, celebrity news and even disasters. Attackers upload malicious posts in the season of special events and disasters. They will upload malicious posts which are related to these events and misguide users to click those links. Users who click the links by mistake act as an adversary to the attacker because the malicious posts would automatically re- posts the malicious contents such as links, images or videos on the user profile. Another popular version of this attack results in user profiles to "like" a Facebook page without their knowledge. In some cases the, spammed posts will lead the users to survey sites which will result in cyber criminals getting profit.

Social network sites provide limited mechanisms to limit the exposure of user profile data to applications. In the case of Facebook, for example, takes an all-or-nothing approach. When users visit an application for the first time, they must give permission to allow that application to access all required profile data. This single choice is to not use or visit the application at all. However, even this does not guarantee any genuine safety.

In this work we develop a system of efficient categorization technique for identifying whether a post generated by a third party application is malicious or not. Detecting malicious URLs is now an essential task in network security intelligence. To maintain efficiency of web security, these malicious URLs have to be detected, identified as well as their corresponding links should be found out. Hence users get protected from it and effectiveness of network security gets increased.

The malicious users can upload a content he wants to spread. The content that contains malicious data is posted to other users wall under a different form. The user mistakes the posts for a real content and clicks the post, which will take him to another page. Thus the malicious user can benefit from this process. In order to get the attention of the user, the malicious user will include keywords or description of

pages that will be of interest to the user. These can be adult content or free downloading sites.

2. Related Works

Using text analytics to detect suspicious user in social media presents an important challenge. There are various methods to detect the meaning of expressions; many works have been done in this context showing several techniques for text analytics.

Sazzadur Rahman [1] has developed FRAppE, an accurate classifier for detecting malicious Facebook applications. Most interestingly, he highlighted the emergence of app-nets—large groups of tightly connected applications that promote each other.

Moreover, Lin et al. [2] was interested in determining the events that are of interests to social networks' users based on their texts data. In this study they collected information from the internet, online communities, and social networks

Sakaki et al. [3] analyze the real-time interaction of micro blogging events especially on Twitter. In their opinion the user may be considered as a sensor to monitor tweets posted recently and to detect different events.

Justin Ma et al. have demonstrated the potential of a classifier based on suspicious URLs [4]. They train their dataset on properties such as host-name length, overall URL length, and the count of the sub domain separating character (.). Combining these lexical features with host information (e.g. DNS registry info), the researchers report an accuracy rate of over 95%.

Gao et al. presented an initial study to quantify and characterize spam campaigns launched using accounts on Facebook [5]. They studied a large anonymized dataset of 187 million asynchronous —wall" messages between Facebook users, and used a set of automated techniques to detect and characterize coordinated spam campaigns. Authors detected roughly 200,000 malicious wall posts with embedded URLs, originating from more than 57,000 user accounts. Following up their work,

Gao et al. presented an online spam filtering system that could be deployed as a component of the OSN platform to inspect messages generated by users in real time [6]. Their approach focused on reconstructing spam messages into campaigns for classification rather than examining each post individually.

3. Problem Statement

Online social networks are widely use these days for the purpose of communication. Users can share more type of information among friends. But there exist some social network users who misuse the features of these social networks and promote the spreading of malicious content. They do this by uploading the malicious post in other user page. These contents spread at a fast rate. There is no proper mechanism to detect these malicious posts immediately and remove it effectively.

4. Proposed Approach

There exists a wide range of malicious content on OSNs today. These include phishing, advertising campaigns, content originating from compromised profiles, artificial reputation gained through fake likes, etc. We do not intend to address all such attacks. We focus our analysis on identifying text posts, malicious URL and creating automated means to detect such posts in real time, without looking at the landing pages of the URLs.

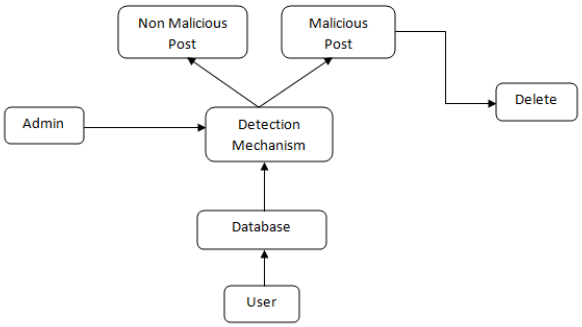

Figure 1: Our Proposed Approach

A. Data Collection and Text Corpus

Text corpus is a huge and structured set of texts posted in the social media, and different techniques can be employed in this step. In this stage we use dataset [9] of Facebook. It's very rich of data users' text posts. and URLs.

B. Corpus Processing

This stage consists to remove stop words and stemming. In computing, stop words are words which are filtered out prior to, or after, processing of natural language data (text). To simplify the study we have to eliminate stop words3 that contains no useful information, as stop word remove stemming [10] can simplify the processing and reduce errors.

C. Classification Process Using Similarity

The system takes malicious post and comments and that are detected by admin using keywords which are there in the admin section. The idea of this approach is to analyze sentences posted by users in social media. System decomposes each post in terms and compares them automatically to suspicious terms. The admin will search for the malicious content on user post and delete that post from the user page and then send an alert to the particular user. The admin can also detect the unwanted comments which appear in the user wall posted by user friends. Based on the request by the user the admin will remove the requested post if it is fount malicious.

If a sentence contains two terms (suspicious words) which presents similarity with the terms of our database we classify as suspicious post. The figure below (refer to figure 2) shows an example of detecting of suspicious post using similarity processing.

We also consider posts which contains at least one URL. If a URL contains any terms which presents similarity with the terms of our database we classify as malicious URLs. Then

admin will delete the URL from user profile and send an alert to particular user. So the admin can prevent a user from uploading malicious post and can also remove the malicious post send by the user friend.

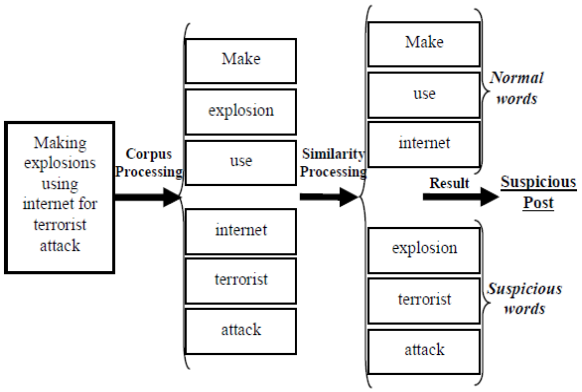

Figure 2: Example of Detection of Suspicious Post Using Similarity Processing

We consider this example ―Making explosions using internet for terrorist attack". After Corpus Processing step, we tested our system using this sentence and we detected three suspicious words which are: explosion, terrorist and attack.

5. Conclusion

The advances in digital and multimedia technology are significantly impacting human behaviors and social interactions. The main idea of our global research project is to develop an automatic system for detecting suspicious profiles in the social media, through which we can uncover suspicious behavior and interests of users as well. The purpose of our approach is to decompose each post in terms and compare them automatically to predefined suspicious terms database by using similarity distance calculation.

In this paper, we have focused to present a system for detecting suspicious posts in social network using similarity approach in text analysis. Our approach is based on similarity with comparing social network seized posts with a suspicious predefined database.

For future work, we plan to improve the system in term of execution time, developing automated classification and using other knowledge resources in order to improve the precision rates, the semantic of exchanged information will be used to identify more significant suspicious profiles.

6. Acknowledgment

I take this opportunity to thank Prof. John Prakash, Prof. P V Bhat & Prof. Dr Nagesh H R, for their valuable guidance and for providing all the necessary support to accomplish this research. I would like to extend my gratitude towards our beloved Principal G L Eshwar Prasad for his great support.

References

[1] Sazzadur Rahman, Ting-Kai Huang, Harsha V. Madhyastha, and Michalis Faloutsos, ―Detecting Malicious Facebook Applications", IEEE/ACM transactions on networking 2015.
[2] C. Lin, B. Zhao, Q. Mei, and J. Han. Pet: ―a statistical model for popular events tracking in social communities". In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010.
[3] T. Sakaki, M. Okazaki, and Y. Matsuo."Earthquake shakes twitter users: realtime event detection by social sensors". In Proceedings of the 19th international conference on World wide web. ACM, 2010.
[4] Ma, Justin, Lawrence K. Saul, Stefan Savage, Geoffrey M. Voelker. "Beyond Blacklists: Learning to Detect Malicious Web Sites from SuspiciousURLs."http://www.csberkeley.edu/_jtma/papers/beyondbl-kdd2009.pdf.
[5] KAUFMAN L., ROUSSEEUW P. J., ―Finding groups in data: An introduction to cluster analysis", WILEYInterscience, 1990.
[6] Vincent Levorato, Thanh Van Le , Michel Lamure, and Marc Bui ―Distance de compression et classification prétopologique ".
[7] M. S. Rahman, T.-K. Huang, H. V. Madhyastha, and M. Faloutsos, ―Efficient and scalable socware detection in online social networks," in *Proc. USENIX Security*, 2012, p. 32.
[8] S.Y Bhat and M. Abulaish,"Community-Based Features for Identifying Spammers in Online Social Networks", ACM,pp. 100-107Aug 25-28 2013.
[9] R. Li, S. Wang, H. Deng, R. Wang and K. C.-C.Chang, "Towards social user profiling: unified and discriminative influence model for inferring home locations," in KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, USA, 2012.
[10] M. F. Porter. An algorithm for suffix stripping. Program, 14(3):130–137, 1980.
[11] Fabr´ıcio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virg´ılio Almeida" Detecting Spammers on Twitter", Computer Science Department, Universidade Federal de Minas Gerais Belo Horizonte, Brazil.
[12] B. Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith."From tweets to polls: Linking text sentiment to public opinion time series". In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media 2010.