www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 10 Oct. 2016, Page No. 18671-18674

Preventing Social Sites from Publishing Malicious Content

Deepak Ranjan, Dr. Tripti Arjariya
Department of Computer Science & Engineering
Bhabha Engineering Research Institute
Bhopal, India
dpkranjan3@gmail.com
Department of Computer Science & Engineering
Bhabha Engineering Research Institute
Bhopal, India
tripti.beri@gmail.com

Abstract— *The World Wide Web has become an inseparable part of millions of people who use online services e.g. online banking, online shopping, social networking, e-commerce, and store and manage user sensitive information, etc. In fact, it is a popular tool for any class of user over the Internet. Rich Web based applications are available over the World Wide Web to provide such types of services. At the same time, the Web has become an important means for people to interact with each other and do business. This is the positive side of this technology. Because the Web can also become a most dangerous place. Due to popularity of World Wide Web which has also attracted intruders and attackers to harm network services. These intruders abuse the users and internet by performing illegitimate activity for their financial profit. These Web pages that contain such types of malicious code are called as malicious Web pages.*

Malicious websites steals the valuable information of the visitors and infect their system for further attacks. Various methodologies are proposed to detect the malicious websites based on features like web contents, HTML codes, session information, and dynamic behaviours.In this paper we are proposed a preventing technique through which we can prevent our web or social sites from publishing malicious content by the intruders.we can check the content and than decide to do published or to do block

.

Keywords— *Malicious Websites, Detection, social sites, preventing techniques.*

I.INTRODUCTION

Malicious Web content [1] has become the primary tool used by attackers to perform attacks on the Internet. In 2007, N. Provos et al. found more than three million URLs that launched drive-by-download attacks. In particular, attacks that target Web clients have become pervasive. According to B. Liang et al. 29 of 90 Websites contained malicious code. According to D. Canali et al. attackers frequently use drive-by-download exploits to compromise a large number of users. To perform a drive-by-download attack, the attacker first craft malicious client-side scripting code typically written in JavaScript that targets vulnerability in a Web browser or in one of the browser's plug-ins. This code is injected into compromised Websites or is simply hosted on a server under the control of the criminals. When victim visits a malicious Web page, the malicious code is executed and the victim's browser is compromised for future attacks. As a result the victim's computer is typically infected with malware.

Drive-By Downloads Attacks

A drive-by-download attack is a malware / virus / shell code delivery technique that is activated simply because the user visited a Website. Drive-by-download attacks occur when a visitor navigates to a site that injects malware onto the victim's PC. A drive-by download can be initiated by simply visiting a Web site or viewing an HTML E-mail message. Basically, these attacks are usually downloaded and run in the background in a manner that is invisible to the user. Drive-by downloads continue to be a major security issue online. In April 2007, researchers at Google discovered hundreds of thousands of Web pages that initiated drive-by downloads. One in ten pages was found to be suspect. Sophos researchers in 2008 reported that they were discovering more than 6,000 new infected Web pages every day, or about one every 14 seconds.

Phishing Attacks

DOI: 10.18535/ijecs/v5i10.55

Phishing is a fraudulent attempt, usually made through E-mail, to steal your personal information, appearing to come from legitimate enterprises (e.g. your university, your Internet service provider, your bank). These messages usually direct you to a spoofed Website or otherwise get you to provide your private information (e.g. password, credit card or other account updates). The attackers then use this private information to commit identity theft. Phishing E-mails will always tell you to click a link that takes you to a site where your personal information is requested. Legitimate organizations would never request this information of you via E-mail.

II. LITERATURE REVIEW

Seminal line of work on content-based anti-spam algorithms has been done by Fetterly et al.. In which they propose that web spam pages can be identified through statistical analysis. Since spam pages are usually automatically generated, using phrase stitching and weaving techniques and aren’t intended for human visitors, they exhibit anomalous properties. Researchers found that the URLs [2] of spam pages have exceptional number of dots, dashes, digits and length. They report that 80 of the 100 longest discovered host names refer to adult websites, while 11 refer to financial-credit-related websites[4]. They also show that pages themselves have a duplicating nature – most spam pages[3] that reside on the same host have very low word count variance. Another interesting observation is that the spam pages’ content changes very rapidly. Specifically, they studied average amount of week-to-week changes of all the web pages on a given host and found that the most volatile spam hosts can be detected with 97.2% based only on this feature. All the proposed features can be found in the paper[6] . In their other work they studied content duplication and found that the largest clusters with a duplicate content are spam. To find such clusters and duplicate content they apply shingling method based on Rabin fingerprints. Specifically, they first fingerprint each of n words on a page using a primitive polynomial PA, second they fingerprint each token from the first step with a different primitive polynomial PB using prefix deletion and extension transformations, third they apply m different fingerprinting functions to each string from the second stage and retain the smallest of the n resulting values for each of the m fingerprinting functions[7]. Finally, the document is represented as a bag of m fingerprints and clustering is performed by taking the transitive closure of the near-duplicate relationship. They also mined the list of popular phrases by sorting (i, s, d) triplets lexicographically and taking sufficiently

long runs of triples with matching i and s values. Based on this study they conclude that starting from the 36th position one can observe phrases that are evidence of machine-generated content[8]. These phrases can be used as an additional input, parallel to common spam words, for a ”bag of word”-based spam classifier.

In which they continue their analysis and provide a handful of other content-based features. Finally, all these features are blended in a classification model within C4.5, boosting, and bagging frameworks. They report 86.2% true positive and 97.8% true negative rates for a boosting of ten C4.5 trees. Recent work describes a thorough study on how various features and machine learning models contribute to the quality of a web spam detection algorithm. The authors achieved superior classification results using state-of the- art learning models, Logit Boost and Random Forest, and only cheap-to-compute content features. They also showed that computationally demanding and global features, for instance PageRank[5], yield only negligible additional increase in quality. Therefore, the authors claim that more careful and appropriate choice of a machine learning model is very important.

III RELATED WORK

Nowadays most people uses internet for various purposes such as online shopping like purchasing or selling products, chat with friends, sending mail. Internet users now spend more time on social networking sites Information can spread very fast and easily within the social media networks. Social media systems depend on users for content contribution and sharing. Facebook had over 1.3 billion active users as of June 2014. there are over 1.3 billion (the number is keep growing) pages from various categories, such as company, product/service, musician/band, local business, politician, government, actor/director, artist, athlete, author, book, health, beauty, movie, cars, clothing, community. Fans not only can see information submitted by the page, but also can post comments, photos and videos to the page.

Xin Jin et. al proposed SocialSpamGuard as spam detection System for Social Media Networks security. Due to the huge amount of posts (over billions) on social media, manually checking every post to pick up the spams is impossible. scalable active learning approach proposed to manually verify as many spams as possible This system has several benefits automatically harvesting spam activities in social network, Introducing both image and text content features and social network features to indicate spam activities Integrating with our GAD clustering algorithm to handle large scale data and Introducing a scalable active learning approach to identify existing spams with limited human efforts, and perform online active learning to detect spams in real-time.

DOI: 10.18535/ijecs/v5i10.55

Kurt Thomas et. al proposed a systematic approach for detecting large-scale attacks on Twitter that we leverage to identify victims of compromise, track how compromise spreads within the social network and evaluate how criminals ultimately realize a profit from hijacked credentials. Criminals succeed in hijacking accounts from users around the globe, irrespective of user understandings. Promising, casual, and core users with hundreds to thousands of followers all fall victim to attacks. At the duration of 1 day in assumed dataset correlate with 21% of victims never returning to twitter after the service wrests control of a victim's account from criminals. Furthermore, 57% of victims lose friends post-compromise in response to spam the victim's account send.

IV PROBLEM DEFINITION

Malicious content detection and work on it has been already done with some sort of area, there are some restriction has been done on posting some unwanted keywords and also repeated sentences in bulk when user try to propagate it with the help of its own id or else it makes number of id using same ip or location or use same metadata in the multiple id. User is still able to propagate with the same sort of content with the help of images or using some hamming images, with the help of images where its textual content is unable to determine by textual mechanism to detect malicious content. Attackers uses image and embedded text in to image and post on the social site, one the user clicks on images the unwanted malicious code is executed . now the attackers uses variety of images like plain image, gray scale images, crop images and also they use multimedia files like videos, xml files, images, etc.

V PROPOSED WORK

For preventing websites from malicious content we proposed content filter techniques. In which some of the points are describe below:

1. *Spamming video*

   Spammers put their spam content into the video. They embed text such as advertisement text in the videos and post these videos to social sites. We can convert these videos into images for detecting malicious content.

2. *Text embedded into image*

   Photos and other graphical images add interest to web pages and printed materials with a minimum of effort. Embedding pictures into your written text is quick and easy, although the process differs greatly depending on the application you are using. It is possible to embed pictures and other graphical images in Word, PowerPoint, Word press and HTML[10].OCR technique will be used here to detect same text in multiple images or multiple text posted by user using multiple ID.

3. *Feature Extraction of image*

   In feature extraction[9] we can further examine the image because user changes its motion in dimensional way or mostly crop its image and we can do de-steganography check. Steganography is the art of hiding the fact that communication is taking place, by hiding information in other information. Many different carrier file formats can be used, but digital images are the most popular because of their frequency on the Internet. For hiding secret information in images, there exists a large variety of stenographic techniques [11] some are more complex than others and all of them have respective strong and weak points. Different applications have different requirements of the steganography technique used.

VI. PROPOSED METHODOLOGY:

Our Steps or Algorithm Steps will follow:

Step 1: Monitor the content which has been posted on websites by the users.

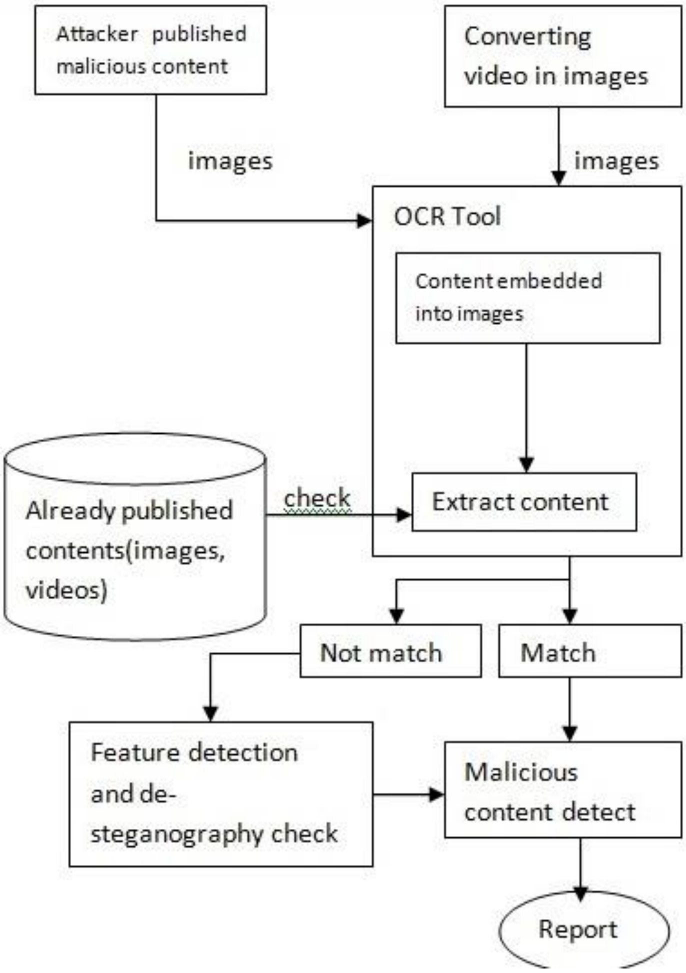
Step 2: If text embedded image is coming than we perform Textual detection from the Image: eliminate of image which is Matched as malicious text with the help of text extracted by OCR tool[10], in this step we are going to use OCR mechanism to process text extraction from the image which is being uploaded by the user.

Step 3:If text embedded video is coming than we first convert the video into multiple image frames and than perform text extraction from OCR tool. If the malicious content is detected than its report to the administration to inform that content is malicious.

Step 4: In this step we are going to work on the feature of spamming image based on image property, its color contribution and then we convert all images into grayscale for reducing the noise overhead and then going to perform the Bayesian algorithm to detect or match the already available image or its related activity. Bayesian algorithm is very efficient algorithm to extract and match the image using its content. And also we will perform anti steganography technique[12] on the image using some common key so that we can detect the unwanted encrypted image which often pass via social network to convey a message, so that our system can be transparent while using image related work on social media.

Step 5: We will use step 2,3,4 where the malicious image and its user with its id can be detected and can be further taken into spammer consideration on the social media.

Step 6: Based on the work we can block the user and can be notify to the administrator.


Attacker published malicious content
Converting video in images
images
images
OCR Tool
Content embedded into images
Extract content
Already published contents(images, videos)
check
Not match
Match
Feature detection and de-steganography check
Malicious content detect
Report

PROPOSED WORK ARCHITECTURE

VII CONCLUSION:

On analysing complete scenario regarding the malicious content posting through the text and on using the images by making they sense to seem a different images than already available images in the present system, and also its uses a text embedded videos.

So here we are taking multiple techniques on which we are going to carry our research by using content filtering concept over the social media websites and network to avoid posting malicious content or to recognize the fake account which assume to get a fake popularity among the social media by using the images (spamming content) of others profile or from other social media available on to the web world.

REFERENCES

[01] Haixu Xi and Hongjin Zhu "Data Mining Methods for New Feature of Malicious Program" , International Journal of Hybrid Information Technology, Vol.9, No.3 (2016), pp. 171-178.

[02] D. Canali, M. Cova, C. Kruegel and G. Vigna, "Prophiler: A fast filter for the large-scale detection of malicious Web pages", Proceedings of the 20th International Conference on World Wide Web (WWW), **(2011)**; Hyderabad, India.

[03] B. Eshete, A. Villafiorita and K. Weldemariam, "BINSPECT: Holistic Analysis and Detection of Malicious Web Pages", Proceedings of the 8th International ICST Conference, SecureComm, **(2012)**; Padua, Italy.

[04] W. Zhang, Y. X. Ding, Y. Tang and B. Zhao, "Malicious web page detection based on online learning algorithm", Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), **(2011)**; Guilin.

[05] V. L. Le, I. Welch, X. Y. Gao and P. Komisarczuk, "Two-Stage Classification Model to Detect Malicious Web Pages", Proceedings of the International Conference on Advanced Information Networking and Application (AINA), **(2011)**; Biopolis.

[06] M. Cova, C. Kruegel and G. Vigna, "Detection and Analysis of Drive-by-Download Attacks and Malicious JavaScript Code", Proceedings of the International World Wide Web Conference Committee (IW3C2), WWW, **(2010)**; Raleigh, North Carolina, USA.

[07] R. B. Basnet and A. H. Sung, "Learning to Detect Phishing Webpages", Journal of Internet Services and Information Security (JISIS), vol. 4, no. 3, **(2014)**, pp. 21-39.

[08] R. B. Basnet and A. H. Sung, "Classifying Phishing Emails Using Confidence-Weighted Linear Classifiers", Proceedings of the International Conference on Information Security and Artificial Intelligence (ISAI), **(2010)**.

[09] George forman,Evan Kirshenbaum "Extremely fast feature extraction for classification and indexing," in LabsHp.

[10] Julinda Gllavata1, Ralph Ewerth1and Bernd Freisleben1,2"A robust algorithm for text detection in image" University of Marburg.

[11] Babloo Saha and Shuchi Sharma, "Steganographic Techniques of Data Hiding using Digital Images" in Defence Science Journal, Vol. 62, No. 1, January 2012, pp. 11-18, DOI: 10.14429/dsj.62.1436.

[12] Clark F. Olson and Daniel P. Huttenlocher , "Automatic Target Recognition byMatching Oriented Edge Pixels" *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 6, NO. 1, JANUARY 1997.*