International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 6 | ISSN : 2456-3307

FACT - Towards Automatic Real Time Identification of Malicious Posts on Facebook

Kanagalakshmi V, Rubygnanaselvam
Bishop Appasamy College of Arts and Science, Race course Road, Coimbatore, Tamilnadu, India

ABSTRACT

Online Social Networks (OSNs) witness a rise in user activity whenever a news-making event takes place. Cyber criminals exploit this spur in user-engagement levels to spread malicious content that compromises system reputation, causes financial losses and degrades user experience. In this paper to detect the malicious contents in the Facebook post using the annotation based approach and compared with the existing WebOfTrust (WOT) method. Social malware posts typically include at least one embedded URL link, since without such a link. The posts cannot lure and hurt users or propagate virally. This approach is trying to detect such posts on the walls and news feeds of Facebook users, and alert users exposed to social malware so that they do not click through on the URLs included in the posts. Final part contains the comparison between the WOT and Fact model on detecting malicious Facebook posts.This model can be used in FACT model to identify the malicious post on eltime,

Keywords: Online Social Networks, WebOfTrus, FACT Model, TDL, WOT

I. INTRODUCTION

The Online Social Networks (OSNs) are new battleground for cybercrime, which provides a fertile and explored environment for dissemination of malware. Moving beyond spam email, the distribution of malware on OSNs takes the form of postings and communications between friends. We use the term social malware to describe parasitic or damaging behavior including identity theft, distribution of malicious URLs, spam, and malicious apps that utilizes OSNs. The use of posts from friends adds a powerful element in the propagation of social malware: it comes implicitly with the endorsement of a friend who allegedly posts the information. This new social dimension adds to the challenges in fighting web-based crime such as the techniques employed by hackers are constantly evolving, and the general public is uninformed, gullible, and easily enticed into visiting suspicious websites or installing apps with the lure of false rewards. Social malware also enables cyber-crime, with several Facebook scams resulting in loss of real money for users and malicious.

Users are enticed into visiting suspicious websites or installing apps with the lure of false rewards, and they unwittingly send the post to their friends, thus enabling a viral spreading. This is exactly where the power of social malware lies: posts come with the implicit endorsement of the sending friend. Beyond this being a nuisance, social malware also enables cyber-crime, with several Facebook scams resulting in loss of real money for users. Cyber criminals have increased their focus on India. According to Kaspersky, an anti-virus solutions company, 7 per cent of all 45 lakh botnet computers that became victims globally are in India. Though there are no official studies, there were few cases from the country until last year. It is very difficult to detect this virus, Kaspersky says. Those who have developed the virus, TDL, Version 4 of which was launched recently, have done a clever job. “It hides in places that the security software rarely looks in. The botnet is controlled using custom-made encryption”. Locating in some part of the world, bot masters get control of data from the infected computers. These infected machines become slaves to attack other computers.

According to the recently released Microsoft's Security Intelligence Report version 9, Indian cybercrooks are increasingly devising techniques to create malicious and computer networks to target and compromising systems, circulate spam e-mails and viruses to other computers through them. As per the report, India is on 25th position in terms of bot infections found and eliminated during the quarter ended-June 2010. The nation accounted for 38,954 computers with bots removed in the second quarter of 2010 compared to 37,895 computers in Q1-2010. Commenting on this matter, Microsoft India Chief Security Officer, Sanjay Bahl stated that it was apparent that the cybercrooks worked hard to sustain and increase them for monetary profits. If it was an ordinary customer then the financial and information loss may not of a great amount, but the loss would be more severe if it was at the organizations or government levels, as reported by economic times on October 14, 2010.

Graham Titterington, Analyst at the research firm Ovum stated that it was clear that the development of the botnet was a main concern. Besides, considerable increase in infections globally, statistics from this year's report showed that cybercrooks were using more complicated methods like malicious to further target prospective victims.

Characterization of malicious content generated in face book events making news.our data set 4,4 million public post in face book literature. Then identify the malicious content is based on like, dislike and angry operation alo perform this paper.

II. RELATED WORK

1. Detecting Social Malware in OSN

This thesis proposes a design and implementation of detecting the malicious posts from the Facebook, which is specifically focus on protecting Facebook users form social malware. Prior solutions for detecting spam and malware on OSNs rely on information obtained either by crawling the URLs included in posts or by performing DNS resolution on these URLs. In contrast, our social malware classifier relies solely on the social context associated with each post Note that this approach means that we do not even resolve shortened URLs into the full URLs that they represent. This approach maximizes the rate at which we can classify posts, thus reducing the cost of resources required to support a given population of users. Machine learning based classification is used to classify the data based on the feature that are readily available from the observed posts. Social malware is a new kind of malware which is significantly different than traditional email spam or web-based malware. First, URL blacklists cannot detect social malware effectively. These blacklists identify only 3% of the malicious posts. The inability of website blacklists to identify social malware is partly due to the fact that a significant fraction of social malware is hosted on popular blogging domains and on Facebook itself.

2. Facebook Terminology

Facebook is the largest online social network today with over 900 million registered users, roughly half of whom visit the site daily. Here, we discuss some standard Facebook terminology relevant to our work.

Post: a post represents the basic unit of information shared on Facebook. Typical posts either contain only text (status updates), a URL with an associated text description, or a photo/album shared by a user. In our work, we focus on posts that contain URLs.

Wall: a Facebook user's wall is a page where friends of the user can post messages to the user. Such messages are called wall posts. Other than to the user herself, posts on a user's wall are visible to other users on Facebook determined by the user's privacy settings. Typically a user's wall is made visible to the user's friends, and in some cases to friends of friends.

News feed: a Facebook user's news feed page is a summary of the social activity of the user's friends on Facebook. For example, a user's news feed contains posts that one of the user's friends may have shared with all of her friends. Facebook continually updates the news feed of every user and the content of a user's news feed depends on when it is queried.

Application: Facebook allows third-party developers to create their own applications that Facebook users can add. Every time a user visits an application's page on Facebook, Facebook dynamically loads the content of the application from a URL, called the canvas URL, pointing to the application server provided by the application's developer. Since content of an application is dynamically loaded every time a user visits the application's page on Facebook, the application developer enjoys great control over content shown in the application page. The Facebook platform uses

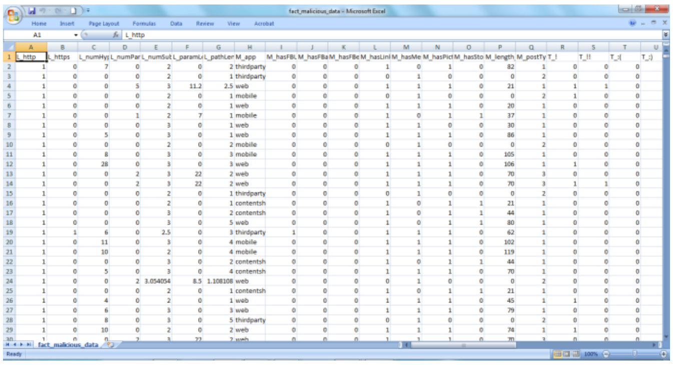
OAuth 2.0 for user authentication, application authorization and application authentication. Here, application authorization ensures that the users grant precise data and capabilities to the applications they choose to add, and application authentication ensures that a user grants access to her data to the correct application.

III. METHODOLOGY

This section describes the proposed model known as FACT model, in which the reputation the post is estimated based on the annotations such as Like, Dislike, Haha, and Angry. A Facebook message which receives more dislikes and angry compared to the overall feedback is referred to as probable malicious content. Hence, only the probable malicious contents only sent to estimate the reputation using WOT services. Therefore, this approach is straightforward and quickly reduces the computation time.

1. FACT Architecture

In this case, we conceptualize and implement an architecture consisting of two aspects referred as feature extraction and classification. The feature extraction model contains forty four features observed from the Facebook messages such as number of words, characters, URLs, hashtags, sentence length, like, dislike, wow, angry and so on. The following Figure 3.3 shows the malicious dataset for FACT model. The classification task is performed through random forest classifier.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	L_http	L_https	L_numHyp	L_numPar	L_numSub	L_paramL	L_pathLen	M_app	M_hasFBL	M_hasFBa	M_hasFBe	M_hasLinl	M_hasMe	M_hasPict	M_hasSto	M_length	M_postTy	T_!	T_!!	T_:(	T_:)
2	1	0	7	0	2	0	2	thirdparty	0	0	0	1	0	1	0	82	1	0	0	0	
3	1	0	0	0	2	0	1	thirdparty	0	0	0	0	1	0	0	0	2	0	0	0	
4	1	0	0	5	3	11.2	2.5	web	0	0	0	1	1	1	0	21	1	1	1	0	
5	1	0	0	0	2	0	1	mobile	0	0	0	0	1	0	0	0	2	1	0	0	
6	1	0	0	0	2	0	1	web	0	0	0	1	1	1	0	20	1	0	0	0	
7	1	0	0	1	2	7	1	mobile	0	0	0	1	0	1	1	37	1	0	0	0	
8	1	0	0	0	3	0	1	web	0	0	0	1	1	0	0	30	1	0	0	0	
9	1	0	5	0	3	0	1	web	0	0	0	1	1	1	0	86	1	0	0	0	
10	1	0	0	0	2	0	2	mobile	0	0	0	0	1	0	0	0	2	0	0	0	
11	1	0	8	0	3	0	3	mobile	0	0	0	1	1	1	0	105	1	0	0	0	
12	1	0	28	0	3	0	3	web	0	0	0	1	1	1	0	106	1	1	0	0	
13	1	0	0	2	3	22	2	web	0	0	0	1	1	1	0	70	3	0	0	0	
14	1	0	0	2	3	22	2	web	0	0	0	1	1	1	0	70	3	1	1	0	
15	1	0	0	0	2	0	1	thirdparty	0	0	0	0	1	0	0	0	2	0	0	0	
16	1	0	0	0	2	0	1	contentsh	0	0	0	1	0	1	1	21	1	0	0	0	
17	1	0	0	0	3	0	2	contentsh	0	0	0	1	0	1	1	44	1	0	0	0	
18	1	0	0	0	3	0	5	web	0	0	0	1	0	1	1	80	1	0	0	0	
19	1	1	6	0	2.5	0	3	thirdparty	1	0	0	1	1	1	0	62	1	0	0	0	
20	1	0	11	0	3	0	4	mobile	0	0	0	1	1	1	0	102	1	0	0	0	
21	1	0	10	0	2	0	4	mobile	0	0	0	1	1	1	0	119	1	0	0	0	
22	1	0	0	0	3	0	2	contentsh	0	0	0	1	0	1	1	44	1	0	0	0	
23	1	0	5	0	3	0	4	contentsh	0	0	0	1	1	1	0	70	1	0	0	0	
24	1	0	0	2	3.054054	8.5	1.108108	web	0	0	0	0	1	0	0	0	2	0	0	0	
25	1	0	0	0	2	0	1	contentsh	0	0	0	1	0	1	1	21	1	0	0	0	
26	1	0	4	0	2	0	1	web	0	0	0	1	1	1	0	45	1	1	0	0	
27	1	0	6	0	3	0	3	web	0	0	0	1	1	1	0	79	1	0	0	0	
28	1	0	8	0	3	0	5	thirdparty	0	0	0	0	1	0	0	0	2	0	0	0	
29	1	0	10	0	2	0	2	web	0	0	0	1	1	1	0	74	1	1	0	0	
30	1	0	0	2	3	22	2	web	0	0	0	1	1	1	0	70	3	0	0	0	

Figure 1. FACT Model – Dataset

Spam keyword score. Presence of spam keywords in a post provides a strong indication that the post is spam. Some examples of such spam keywords are FREE, Hurry, Deal, and Shocked. To compile a list of such keywords that are distinctive to social malware, our intuition is to identify those keywords that 1) occur frequently in social malware posts, and 2) appear with a greater frequency in social malware as compared to their frequency in benign posts.

We compile such a list of keywords by comparing a dataset of manually identified social malware posts with a dataset of posts that contain URLs that match our whitelist. We transform posts in either dataset to a bag of words with their frequency of occurrence. We then compute the likelihood ratio p1=p2 for each keyword where p1 = p(wordjsocialmalwarepost) and p2 = p(wordjbenignpost). The likelihood ratio of a keyword indicates the bias of the keyword appearing more in social malware than in benign posts. In our current implementation, we have found that the use of the 6 keywords with the highest likelihood ratio values among the 100 most frequently occurring keywords in social malware is sufficient to accurately detect social malware.

Thereafter, to classify a URL, our architecture searches all posts that contain the URL for the presence of these spam keywords and computes a spam keyword score as the ratio of the number of occurrences of spam keywords across these posts to the number of posts.

Message similarity. If a post is part of a spam campaign, it usually contains a text message that is similar to the text in other posts containing the same URL (e.g., because users propagate the post by simply sharing it). On the other hand, when different users share the same popular URL, they are likely to include different text descriptions in their posts. Therefore, greater similarity in the text messages across all posts containing a URL portends a higher probability that the URL leads to spam.

To capture this intuition, for each URL, we compute a message similarity score that captures the variance in the text messages across all posts that contain the URL. For each post, it sums the ASCII values of the characters in the text message in the post, and then computes the standard deviation of this sum across all the posts that contain the URL. If the text descriptions in all posts are similar, the standard deviation will be low.

Like and comment count. Facebook users can ‘Like’ any post to indicate their interest or approval. Users can also post comments to follow up on the post, again indicating their interest. Users are unlikely to `Like' posts pointing to social malware or comment on such posts, since they add little value. Therefore, for every URL, it computes counts of the number of Likes and number of comments seen across all posts that contain the URL.

URL obfuscation. Hackers often try to spread malicious links in an obfuscated form, e.g., by shortening it with a URL shortening service such as bit.ly or goo.gl.

However, as we show later in our evaluation, the features that we currently consider yield high classification accuracy in combination

```
ALGOITHM;
For all posts do
        content = (no. of dislike)+(no. of angry)/total emotions
        If content > 50% then
                Post = probable_malicious
End if
End for
For all probable posts do
For all URL domains do
Components = GetComponentFromWOT_API
For all components do
If reputation < 60 and confidence ≥ 10 then
Post = malicious
End if
End for
End for
End for
```

This algorithm mainly tries to estimate the reputation of the Facebook poststhrough annotation of emotions recorded under the post. The dataset contains various features; in particular emotional features such as angry, dislike are the basic symptom to identify the posts as bad emotion. If the bad emotion ratio is greater than 50% then it is suggested as probable malicious content. This approach drastically reduce the number of records in the dataset, which means benign data will be clearly marked as benign and other records as probable malicious content. Further, the probable malicious content only considered to validate the reputation. Therefore, if the reputation score is below 60 marked as malicious and other scores retained as probable malicious. Hence, it reduces the computation time and improthe prediction accuracy

IV. RESULT

1. PERFORMANCE EVALUATION

As mentioned in the previous chapters, the main intention of this thesis is to validate the malicious post through annotation and reputation score. The base model uses WOT to estimate the reputation; whereas the proposed method uses emotions based annotation along with WOT to estimate the reputation. The dataset contains 44 features, and random forest classifier is used to predict the classification accuracy. The combination of above techniques validates around 21387 out of 22434 posts declared as social malware by the WOT classifier are indeed so. Therefore, 95.33% of the social malware identified by WOT's classifier are true positives. On the other hand, the 1047 posts incorrectly classified as social malware constitute less than 0.046% of the over 20 thousand posts in our dataset. Note that, though all of the above techniques could be folded into FACT model itself to help identify social malware, we do not do so because all of these techniques require us to crawl a URL in order to evaluate it; we cannot afford the latency of crawling.
The Figure 4.1 shows the classification accuracy of WOT and FACT model, while predicting the malware from the Facebook posts. The result states that the FACT model finds 84.4% of social malware URLs whereas the WOT model finds 82.1% of social malware.

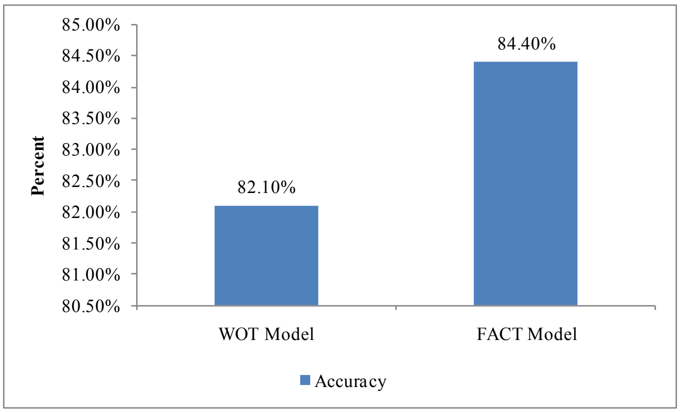


Figure 1. Accuracy of WOT and FACT models

The following Figure 4.2 shows the execution time of the WOT and FACT model to predict the malicious posts among the given dataset. The result states that

WOT model consumes 0.42 seconds. The FACT model consumes 0.35 seconds, which is comparatively lesser time than the WOT model. Therefore, it can be concluded that the proposed FACT model is better than the plain WOT model.

V. CONCULISION

The Online Social Networks (OSNs) has opened up new possibilities for the dissemination of malware. As Facebook is becoming the new web, hackers are expanding their territory to Online Social Networks (OSNs) and spread social malware. Social malware is a new kind of cyber-threat, which requires novel security approaches. Online fraud is an immediate and expensive problem that affects people and business through identity theft, the spread of viruses, and the creation of botnets, all of which are interconnected manifestations of Internet threats. Web of Trust (WOT) is a community powered approach which enables to detect the malicious contents. Similarly, we propose a FACT model, which includes emotions based annotation along with WOT. The result states that the FACT model finds 84.4% of social malware URLs whereas the WOT model finds 82.1% of social malware. Similarly, WOT model consumes 0.42 seconds and FACT model consumes only 0.35 seconds. Therefore, it is concluded that the proposed FACT model is effective in terms of accuracy and time consumption.

VI. REFERENCES

[1]. Alex Wang. Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach. In Data and Applications Security and Privacy XXIV. 2010.

[2]. Andreas Makridakis, Elias Athanasopoulos, Spiros Antonatos, Demetres Antoniades, Sotiris Ioannidis, and Evangelos P. Markatos. Understanding the behaviour of malicious applications in social networks. Netwrk. Mag. of Global Internetwkg., 2010.

[3]. Andrew Besmer, Heather Richter Lipford, Mohamed Shehab, and Gorrell Cheek. Social applications: exploring a more secure framework. In SOUPS, 2009.

[4]. Anestis Karasaridis et.al., "Wide-scale botnet detection and characterization", HotBots'07 Proceedings of the first conference on First Workshop on Hot Topics in Understanding Malicious USENIX Association Berkeley, CA, USA 2007, pp. 7 - 7.

[5]. Anh Le, Athina Markopoulou, and Michalis Faloutsos. Phishdef: Url names say it all. In Infocom, 2010.