An Anti-Image Technique for Sybil Detection in Web Data

Abhishek Gaur
Department of Computer Science & Engineering
Lakshmi Narain College of Technology
Bhopal, India

Ratnesh Kumar Dubey
Asst. Professor
Department of Computer Science & Engineering
Lakshmi Narain College of Technology
Bhopal, India

Vineet Ricchariya, Ph.D
Prof. & Head
Department of Computer Science & Engineering
Lakshmi Narain College of Technology
Bhopal, India

ABSTRACT

Today from security point of view, the increasing demand of network connectivity makes the system insecure. So we need a complement that can cope / prevent the security breaches in system. Unfortunately, in many environments, it may not be feasible to render the computer system immune to all type of intrusions. The motivation behind this project is to develop a complement system i.e. Sybil detection that can prevent all possible breaks-ins. In Sybil attack each node can be preferred to as a separate webpage that the spammer creates, and each of these webpages interlink to each other thus forming a network similar to link farms. The spamming webpages link to other nodes and thus create a huge linked for improving popularity.

In this paper we are introduce anti-image spamming technique for preventing to Sybil attack to attackers .Our objective is to work on to the real time research desired scenario where the work is needed to be done. Sybil attack is one of the scenario where a work has been done to deal with Sybil attack efficiently in some of the field but still there are more has to be done which we want to carry forward with the help of spamming image detection and prevention technique with the help of image anti-spamming technique. In this paper we present result analysis based on the Sybil detected by different technique performed by us.

Keywords

Anti-image spamming, Sybil attack, Social Network, OCR Technique

1. INTRODUCTION

The Synonym Identity or Sybil attack is an attack or a phenomenon where in a trusted system is subverted by duplicate or wrong identities known to one network [1]. A malicious user pretends to be multiple nodes in the system by faking identities. Sybil attack is a black hat SEO manipulation where a spammer takes over the trusted systems of various networks like forums, blogs, social networking sites like Facebook, Twitter etc [5]. They create multiple identities using each one of these networks in order to improve the reputation of the main ID. With this reputation they post 'n' number of ads and posts hoping that they don't get noticed. A spammer may also create Sybil attack by creating a lot of sites that link to each other. These sites may be pure spam blogs or irrelevant pages with low quality content. Using Sybil attack to manipulate the search engine is very rare but still there are spammers who use such tactics to gain traffic to their sites. Search engines are trying to take action towards such attacks. Sybil Attack as of multiple identities for malicious intent named after the famous multiple personality disorder patients "Sybil". This particular attack has been used by spammers to create multiple websites or ids with identical domain names with junk and duplicate content. These pages have no quality content and are created just with the intention to create spam and drive traffic. All these webpages are interlinked to each other in order to boost their search engine traffic. In Sybil attack each node can be preferred to as a separate webpage that the spammer creates, and each of these webpages interlink to each other thus forming a network similar to link farms. The spamming webpages link to other nodes and thus create a huge linked for improving popularity.

Spammers also attack onto the web or social network by putting their spam content into the image. They embed text such as advertisement text in the images and attach these images to emails. Anti-spam filter that analyse content of email cannot detect spam text in image. In order to allow the same kind of image with the multiple ids they make spam image of the original image in different manner and present in the social network system. In this paper we are going to work with mentioned sort of spammers those propagate the Sybil attack using spamming image in peer to peer network.

Type of image spam we are going to perform:

1. Text-only image
2. Sliced image
3. Randomized image
4. Gray images
5. Colour modified image
6. Wild background image

2. LITERATURE REVIEW

One practical limitation of structured peer-to-peer (P2P) networks is that they are frequently subject to Sybil attacks: malicious parties can compromise the network by generating and controlling large numbers of shadow identities. In this paper, we propose an admission control system that mitigates Sybil attacks by adaptively constructing a hierarchy of cooperative peers. The admission control system vets joining nodes via client puzzles. A node wishing to join the network is serially challenged by the nodes from a leaf to the root of the hierarchy. Nodes completing the puzzles of all nodes in the chain are provided a cryptographic proof of the vetted identity. We evaluate our solution and show that an adversary must perform days or weeks of effort to obtain even a small percentage of nodes in small P2P networks, and that this effort increases linearly with the size of the network. We further show that we can place a ceiling on the number of IDs any adversary may obtain by requiring periodic reassertion of the IDs continued validity.

Douceur [6] was the first to consider multiple identity problems in the context of P2P networks. Dubbed the “Sybil” attack, the registration of many new nodes to take control of a system plagues more than just P2P networks. Any distributed system in which an entity can arbitrarily establish identities, is subject to its effects. The designers of the original structured P2P overlays paid little attention to the severity of Sybil attacks; most schemes either neglect to consider it or include limited defenses. For example, in Chord [7] and Pastry [8], the authors assumed that a node’s ID was the hash of its IP address. However, an adversary can simultaneously spoof many IP addresses to quickly obtain a multitude of identities. Additionally, using hashed IP addresses limits access to the network from machines behind NAT boxes. In CAN [9], the authors assumed that nodes pick random IDs when they enter the network. This places trust on all nodes in the system and easily allows an adversary to create many IDs. Many different types of cryptographic solutions to the Sybil attack have been proposed. While the application of cryptography potentially provides a solution, no current method efficiently mitigates the attacks. Because Sybil attacks result from entities misidentifying themselves, requiring all nodes to authenticate with public keys is a one approach to securing these networks. Douceur [6] showed that without the use of a centralized authority [10] that certifies all nodes, it is impossible to prevent this attack. Srivatsa and Liu [11] suggested the use of certificates with limited lifetime issued by the bootstrap entry point that binds a node with a unique ID. This would limit the number of IDs an adversary can obtain during a time period and will depend on the lifetime of the ticket. However, requiring all nodes to obtain a certificate that will bind it with a unique ID is not only expensive but will require either releasing private information or paying an amount of money for the service.

3. RELATED WORK

Already the work with Sybil attack has been done while dealing with the textual content and dealing with the same IP detection for multiple id [2][4], but till the attacker can do the attack using the multiple image upload and also they get affect server and Sybil attacking process in the form of steganography image and propagate them without any issue.

This is still an open area to work on Sybil attack with anti-image spamming policies, as we have already number of anti-image spamming policy to detect and classify spamming images here we can apply the related algorithm and can prepare and framework which can be apply on to the existing system to prevent Sybil attack effectively [12].

Sybil Attack as of multiple identities for malicious intent named after the famous multiple personality disorder patients "Sybil". This particular attack has been used by spammers to create multiple websites or id’s with identical domain names with junk and duplicate content. These pages have no quality content and are created just with the intention to create spam and drive traffic [3]. All these webpages are interlinked to each other in order to boost their search engine traffic. In Sybil attack each node can be preferred to as a separate webpage that the spammer creates, and each of these webpages interlink to each other thus forming a network similar to link farms. The spamming webpages link to other nodes and thus create a huge linked for improving popularity .In order to work with the Sybil profile detection in various conditions we have performed the experiment and the proposed methodology derived and performed by us on image and textual data.

Our work performed the contribution in Sybil detection technique in text as well as image area.

Our Steps or Algorithm Steps will follow:

Step 1: Monitor the image which is being uploaded by the user in any form, either in profile update or sharing scenario.

Step 2: We perform Textual detection from the Image: eliminate of image which is Matched as Sybil text with the help of text extracted by OCR tool [14], 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: In this step we are going to work on the feature of spamming image based on image property [16], its colour contribution and then we convert all images into grey scale for reducing the noise overhead and then going to perform the Bayesian algorithm [13] 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.

Step 4: in this step we will perform anti steganography technique [15] 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 spamming image and its user with its id can be detected and can be further taken into spammer consideration or as Sybil attacker on the social media.

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

OCR TECHNIQUE

OCR is a technique which is used for text detection from the various image format , we are using this technique for finding the spam images which uses spam text contained text ,for this we are using Tess4J API to detect text from the images.

Following is the Algorithm which we have used:

```
File f=new File("G://Dataset");
        File fg[]=f.listFiles();
        int t=fg.length;
        for (int i = 0; i < t; i++) {
File imageFile = new File(fg[i].getAbsoluteFile().toString());
      ITesseract instance = new Tesseract();
      String result="";
      result = instance.doOCR(imageFile);
}
```

Which we are using for the text detection from the image, here reading all the image dataset and finding the text contain on it and storing into the result for further detection.

DE-STEGANOGRAPHIC CHECK

We are using text extraction from image using de-stragenographic technique on image and finding the text if hidden behind the image or media shared by the user.

Following is the algorithm we have used to determine:

```
 int len = extractInteger(image, 0, 0);
   byte b[] = new byte[len];
   for(int i=0; i<len; i++)
      b[i] = extractByte(image, i*8+32, 0);
   message.setText(new String(b));
   return new String(b);
```

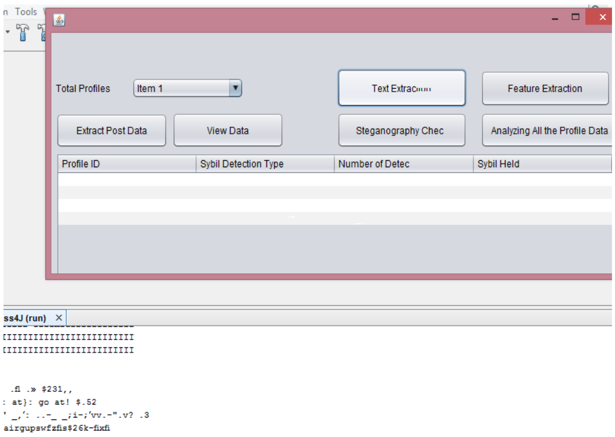
Above code provide us the detected or contained text from the image. Where we provide all the images in an loop and it checks the hidden text behind the image if any user has embedded into it.

FEATURE ANALYSIS

A feature analysis algorithm is used which check the various feature parameter associate with image such as height, width and its pixel count such that similar images can be determine where we are observing 0.02 for 90% matching ,0.05 for average matching, 0.1 for exact matching.

4. EXPERIMENTAL & RESULT ANALYSIS

All the experiments were performed using an i5-2410M CPU @ 2.30 GHz processor and 4 GB of RAM running windows 7. The discussed feature selection algorithms were implemented using language Java. Swing includes graphical user interface (GUI) widgets such as text boxes, buttons, split-panes, and tables. Swing widgets provide more sophisticated GUI components than the earlier Abstract Window Toolkit. Proposed as well as existing algorithms were applied one by one in both the proposed framework from dataset [17].



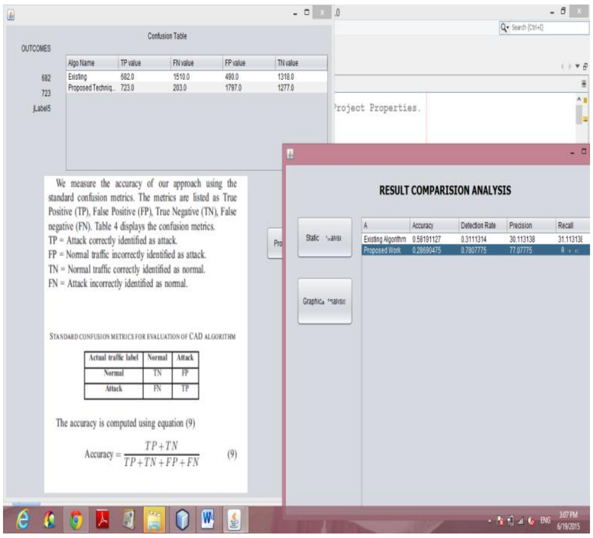
Profile ID	Sybil Detection Type	Number of Detec	Sybil Held

```
IIIIIIIIIIIIIIIIIIIIIIIII
IIIIIIIIIIIIIIIIIIIIIIIII


 .fl .» $231,,
: at}: go at! $.52
' _,': ..-_ _;i-;'vv.-".v? .3
airgupswfzfis$26k-fixfi
```

Figure 1: Proposed Experimental framework

Upon working with the designed framework we have come up with the result using existing dataset and experiment performed using existing and proposed technique on textual and image dataset. We have analysed the result with parameter accuracy, detection rate, precision and recall calculated using our programming and result obtain statically and graphically which are presented below:



Confusion Table

OUTCOMES

682

723

jLabel5

Algo Name	TP value	FN value	FP value	TN value
Existing	682.0	1510.0	490.0	1318.0
Proposed Techniq...	723.0	203.0	1797.0	1277.0

We measure the accuracy of our approach using the standard confusion metrics. The metrics are listed as True Positive (TP), False Positive (FP), True Negative (TN), False negative (FN). Table 4 displays the confusion metrics.
TP = Attack correctly identified as attack.
FP = Normal traffic incorrectly identified as attack.
TN = Normal traffic correctly identified as normal.
FN = Attack incorrectly identified as normal.

STANDARD CONFUSION METRICS FOR EVALUATION OF CAD ALGORITHM

Actual traffic label	Normal	Attack
Normal	TN	FP
Attack	FN	TP

The accuracy is computed using equation (9)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (9)$$

RESULT COMPARISION ANALYSIS

A	Accuracy	Detection Rate	Precision	Recall
Existing Algorithm	0.58191127	0.3111314	30.113138	31.113138
Proposed Work	0.28690475	0.7807775	77.07775	[illegible]

Figure 2: Number of Attribute received after applying outcomes technique

Here in our research we are experimentally going to provide a simulation Sybil system where query will work with different type of conditions related to past behavior of query and ontology database and will help to provide accurate result. Here are the described result screen and detail discussion about the result. Number of Outcomes we have detected after performing outcomes technique on the complete dataset and performing attribute based ranking on individual technique, After applying such technique we have received few dataset to further work on and optimization.

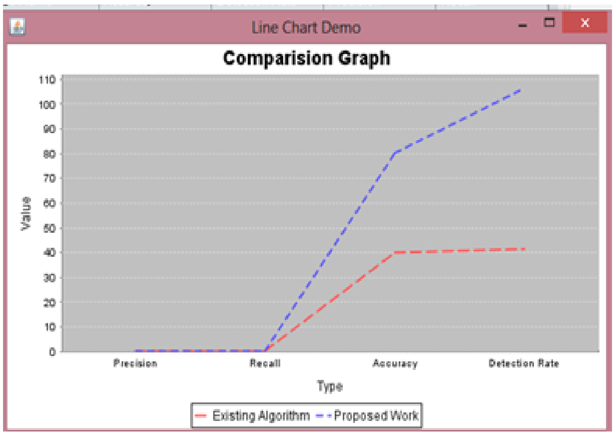


Figure 3: Graph – Dataset retrieved on applying different dataset attribute

Above graph is the line graph which is drawn using Jfree API java [18] which present our statical result into graphical such than it representing the difference between the existing and proposed scheme results.

5. CONCLUSION

In this paper we have shown the work which is defined to work with the Sybil attack on considering image spanning conditions.

We have examined the profile images using various techniques such as OCR, feature extraction and De-steganography technique to verify username and its profile data where it matches as Sybil or not. Our contributions perform and monitor the keywords from textual and image format manner whether it comes or match with the existing profiles or not. Further on based on similar keywords shared by the multiple profiles has been differentiated and taken as Sybil profile users. Our work performs an extra effort when we are working with multimedia image files and thus we are getting better result in the parameter recall, precision using the different technique. Our future work will be to find more algorithm and approach which can work on the Sybil image detection and also we can further work on other multimedia format such as video and others.

6. REFERENCES

[1] Neil Zhenqiang Gong, Student Member, IEEE, Mario Frank, and Prateek Mittal, Member, IEEE ”SybilBelief: A Semi-Supervised Learning Approachfor Structure-Based Sybil Detection” IEEE Transactions On Information Forensics And Security, Vol. 9, No. 6, June 2014.

[2] Bimal Viswanath, Mainack Mondal, Allen Clement, Peter Druschel, Krishna P. Gummadi, Alan Mislove, and Ansley Post, “Exploring the design space of social network-based Sybil defenses”, IEEE 2012.

[3] C. Lesniewski-Laas and M. F. Kaashoek. Whanau: A sybil-proof distributed hash table. In Proc. NSDI’10, San Jose, CA, Apr 2010.

[4] B. Vishwanath, A. Post, K. Gummadi and A. Mislove. An analysis of social network-based Sybil defenses. In Proc. Of ACM SIGCOMM, 2010.

[5] C. Zhang and J. Sun. Privacy and Security for Online Social Networks: Challenges and Opportunities. In IEEE Network, 2010.

[6] J. Douceur. The sybil attack. In Proceedings of the First International Workshop on Peer-to-Peer Systems 200, Cambridge, MA, March 2002.

[7] I. Stoica, R. Morris, D. Karger, M. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In SIGCOMM 2001, August 2001.

[8] A. Rowstron and P. Druschel. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In IFIP/ACM Middleware. Heidelberg, Germany, 2001.

[9] S. Ratnasamy, P. Francis, M. Handley, and R. Karp. A scalable contentaddressable network. In SIGCOMM 2001.

[10] E. Sit and R. Morris. Security considerations for peer-to peer distributed hash table. In 1st International Workshop on Peer-to-Peer Systems, Cambridge, MA, March 2002.

[11] M. Srivatsa and L. Liu. Vulnerabilities and security threats in structured overlay networks: A quantitative analysis. In ACSAC 2004.

[12] AXinwen Fu and Zhen Ling, “One Cell is Enough to Break Tor’s Anonymity” in Internet Measurement Comference, 2009, pp. 29–42.

[13] George forman,Evan Kirshenbaum “Extremely fast feature extraction for classification and indexing,” in LabsHp.

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

[15] 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.

[16] 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*.

[17] http://wis.ewi.tudelft.nl/umap2011/#dataset.

[18] http://www.jfree.org/jfreechart/api/javadoc/org/jfree/chart/package-summary.html.