Sybil Attack detection approach for detection and prevention of social image & textual data

Salil Puri*, Mukesh Kumar 2** and Sitesh Kr. Sinha***
*Department of (CSE), *Student IV Semester, AISECT UNIVERSITY, Bhopal, India*
salilpuri@yahoo.com
*** HOD CSE, College, AISECT UNIVERYSITY, Bhopal*

***Abstract:* Sybil attack and its spamming with the similar type of data populating among the available platforms. Data increasing in social media with an exponential growth in different format of storage such as image, textual and other format of sharing methods available. Various organization and people try to build an self propagating and create multiple identity to process their marketing content. In the view of reaching number of customers and showing the large number of followers to them. This comes to the Sybil identity which exhibit the large number of identity for same person with variants in uniqueness. Processing and finding the textual content similarity and identity is defined in past work which is effective by using NLP (Natural language processing) and other modularity given. Past approach help in learning textual data behavior, its purpose, its impact on the user by content matching, semantic matching or other neural based approach.**

In this research a proposed approach which is combination of identify textual similarity, image similarity using de-steganography, by using OCR approach over the text on image. Thus the technique embeds a hybrid searching of Sybil over textual and multimedia image data. A result analysis is performed with java API, which shows the efficiency of our proposed approach over existing scenario.

Keywords: Sybil Analysis, Image computation, OCR technique, NLP, OSN platform analysis.

INTRODUCTION The popularity of the term social networking web sites has been increased since 1997, and millions of people now are using social networking web sites to communicate with their friends, perform business and many other usages according to the interest of the users. The interest of social networking web sites has been increased and many research papers have been published. Some of them discussed the security issues of social networking, analyzing the privacy and the risks that threat the online social networking web sites.

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. 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.

RELATED WORK

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 [5] 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.

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.

PROPOSED WORK

Our proposed approach consist of various steps which take participation of data usage, data processing and performing identity from the textual and graphical content using sub-algorithms.

Step 1: Loading the complete social media data set downloaded from the available resources and updating in framework.

Step 2: A further performance over the data is drawn using NLP and other image data extraction technique which help in content understanding such as verb, noun and other object identities available in dataset. Image data extraction, feature extraction is performed.

Step 3: In this step we are going to work on the image data and textual data feature extraction which help in computing an score based on which a further data computation and analysis is performed. A classification algorithm which is bayes very efficient algorithm to extract and match the image using its content.

Step 4: in this step De-steganography approach is performed to understand any hidden Sybil content is given in the data. Data simulation , Sybil and negative word understanding thus to find Sybil is executed.

Step 5: We will use step 2, 3, 4 to understand textual spamming content and image hidden spamming content. Thus by obtaining contents, the Sybil identities over matching data is performed over the dataset.

Step 6: Based on the work we can further an action can be taken.

RESULT ANALYSIS

As per the analysis of system , our algorithm take an advantage of proposed algorithm hybrid system which help in utilizing of previous technique , determining detection over the given area.

Statically analysis:

In the table 1 below, an comparison analysis in the term of precision , recall, detection rate and accuracy is derived.

Technique/ Parameters %	Existing Textual Semantic approach	Multimedia based Hybrid approach
Precision	88.65	89.43
Recall	78.4	63.5
Accuracy	91.01	92.65
Detection Rate	89.66	94.21

Table 1: Result analysis using existing and proposed hybrid approach.

Graphical Analysis:

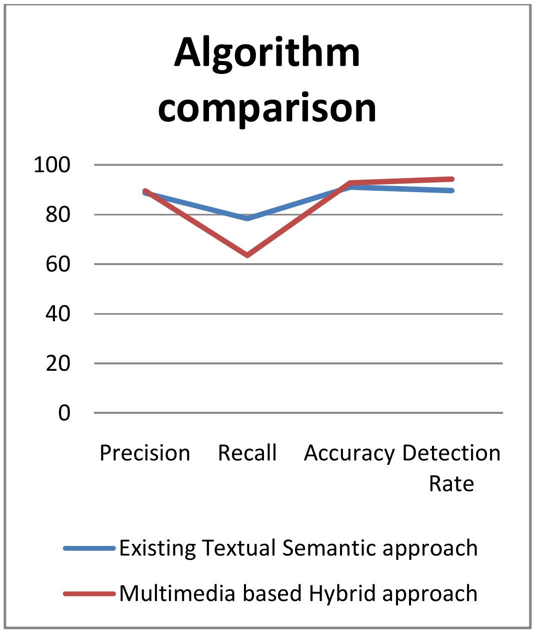

Figure 1: Graphical analysis of result obtained and its comparison

In the figure 1 above, a graphical analysis of given approach is performed which shows the efficiency of proposed approach.

The Proposed textual and graphical analysis shows the proposed technique gives an proper usage advantage over the previous technique. The proposed technique compares in terms of precision, recall, accuracy and detection rate. It exhibit proposed data ratio and perform efficient work.

CONCLUSION

Sybil attack is an important aspects today in various social media and generic platform which talk about the usage of their data . Social media platform help in organic visibility of users content, they help in advertising and reach of correct content to the genuine number of public. It works according to user interest and their application use. Social media platform face a challenge from the different user and organization such as facing a Sybil identity and populating wrong data among the users. It make use of false review, rating , like and other scenario by the fake id created by Sybil . Sybil can be identify by identifying proper given text or other data by user. Identifying textual content, identifying multimedia content is important aspect.

In this research an identification of Sybil and finding attack from the different resources users on the web social is performed. Identification of similar text, similarity content in image multimedia data is performed. This research analyze the past technique and hence an enhancement of their usage, use of OCR, De-Steganography technique is utilized to determine hidden text with in the image. To detect similarity in textual content, same semantic usage is taken for consideration.

Our work is compared with existing technique and found suitable while working with proposed hybrid approach given by us. A Simulation is performed using Java API 8, with Net beans tool. Sample large data set identities were taken to process and found a usage result over the previous and proposed technique. Our final work shows an efficient outcome.

FUTURE WORK

As the discussion is made while working with textual and multi-media data to work towards data processing. Identifying Sybil , similar data textual and from the multiple multimedia image format is done.

A further aspect of working with is given here:

1. A scenario with real time usage of the proposed technique can be taken into consideration, which can help in working with the generated online social media platform.
2. Working with large dataset is also one of the future working direction, which can prove our work accuracy.
3. Working with more format of multimedia data such as extension version of image, Working with video formats.
4. Final version can further be computed with more available technique in the same area of computation.

REFERENCES

[01] Xiaohui Liang and Xiaodong Lin. “Fully Anonymous Profile Matching in Mobile Social Networks”, IEEE areas in communication vol:31 no:9 year 2013.

[02] N. Eagle and A. Pentland, “Social serendipity: mobilizing social software, ”IEEE Pervasive Computing, vol. 4, no. 2, pp. 28–34, 2005.

[03] J. Teng, B. Zhang, X. Li, X. Bai, and D. Xuan, “E-shadow: Lubricating social interaction using mobile phones,” in ICDCS, 2011, pp. 909–918.

[04] B. Han and A. Srinivasan, “Your friends have more friends than you do: identifying influential mobile users through random walks,” in MobiHoc, 2012, pp. 5–14.

[05] Z. Yang, B. Zhang, J. Dai, A. C. Champion, D. Xuan, and D. Li, “E-smalltalker: A distributed mobile system for social networking in physical proximity,” in ICDCS, 2010, pp. 468–477.

[06] R. Lu, X. Lin, X. Liang, and X. Shen, “A secure handshake scheme with symptoms-matching for mhealthcare social network,” ACM MobileNetworks and Applications (MONET), vol. 16, no. 6, pp. 683–694, 2011.

[07] J. Katz, A. Sahai, and B. Waters, “Predicate encryption supporting disjunctions, polynomial equations, and inner products,” in EUROCRYPT, 2008, pp. 146–162.

[08] M. Motani, V. Srinivasan, and P. Nuggehalli, “Peoplenet: engineering a wireless virtual social network,” in MobiCom, 2005, pp. 243–257.

[09] M. Brereton, P. Roe, M. Foth, J. M. Bunker, and L. Buys, “Designing participation in agile ridesharing with mobile social software,” inOZCHI, 2009, pp. 257–260.

[10] E.Bulut and B.Szymanski, “Exploiting friendship relations for efficient routing in delay tolerant mobile social networks,” IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 12, pp. 2254–2265, 2012.

[11] Z. Yang, B. Zhang, J. Dai, A. C. Champion, D. Xuan, and D. Li, “E-smalltalker: A distributed mobile system for social networking inphysical proximity,” in ICDCS, 2010, pp. 468–477.