A Survey: Identity Matching Technique over the Social Media Profile

Monika Saunakiya*, Prof. Damodar Tiwari** and Dr. Shishir K. Shandilya***
*(0112CS15MT07)
*-***Bansal Institute of Science and Technology, Bhopal

Abstract: Social media is a platform where various users share their emotions and feelings every day. Many industries also work towards finding opportunity through social media and promoting their business towards public interest. Behind publishing their content over social media, organizer also performs multiple identity creation to promote their product and its visibility. Various identity formation and working along with their identity is propagating by the users. Various research is also performed to identity those identity and blocking them. In this paper a survey of different approaches which take participate in social multiple identity detection of same user. The paper also shows the drawback of previous approach given by them and further possible solutions to resolve them. The paper includes current author reviews and solutions over identity matching in social media.

Keywords: social identity detection, Sybil matching, data processing, identity prevention, Sybil model.

Introduction

The Synonym Identity or Sybil attack is an attack or a phenomenon where in a trusted system is reveal by redundant or wrong identities known to one network and another [1]. A malicious and attacker user pretends keep to be a number of and redundant large number of nodes in the system by faking identities to act as a Sybil system in Social media sites and networking area site. Sybil attack in a Social media sites and networking area is a black hat SEO(Search oriented website or blog optimization) manipulation where a unwanted data and un-necessary messages or spam content messenger takes over the trusted systems of various networks like informative and connecting forum services, blogger spot services, social media connecting and networking sites like FB(Facebook Social Networking site), Twitter etc[2,4].

They create a number of and redundant large number of 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. An unwanted data and un-necessary messages or spam content massager may also create Sybil attack by creating a lot of sites that link to each other. These sites may be pure unwanted data and un-necessary messages or spam content message blogger spot services or unmatched data content pages with low quality content. Using Sybil attack to manipulate the Search oriented website or blog is very rare but still there are unwanted data and un-necessary messages or spam content messengers who use such tactics to gain traffic to their sites. Search engines are trying to take action towards such attacks. Sybil Attack as of a number of and redundant large number of identities for malicious intent named after the famous a number of and redundant large number of personality disorder patients "Sybil" [7,8]. This particular attack has been used by unwanted data and un-necessary messages or spam content messengers to create a number of and redundant large number of web pages and data's or ids with identical domain names with junk and redundant content. These pages have no quality content and are created just with the intention to create unwanted data and un-necessary messages or spam content message and drive traffic. All these WebPages are interlinked to each other in order to boost their Search oriented website or blog traffic. In Sybil attack each node can be referred to as a separate webpage that the unwanted data and un-necessary messages or spam content messenger creates, and each of these WebPages interlink to each other thus forming a network similar to link farms. The unwanted data and un-necessary messages or spam content messaging WebPages link to other nodes and thus create a huge linked for improving popularity[13].

The rest of the survey paper is organized as Literature survey , problem formulation , overview of proposed solution , conclusion over literature discussion[11].

Literature Review

Sybil identity over the social media is increasing over a long span. Finding multiple identities creating by same person or organization is performed by research authors. In this section previous research performed is discussed.

In this paper [1] sybilvote algorithm is proposed, for the Sybil estimation they have used some algorithm which is based on different formulas and computation of inputs. The computation formulas are Monte-Carlo simulation and more accurate than the existing formula based on the multinomial distribution tail estimate. The approximation formula over detection real user

and fake user with the score values. The computation complexity and Sybil detection with Sybil ratio is compared in the paper. An accurate approximation formula gives the exact value proposition on proposed system.

In this paper [7] author presented Sybil identity detection over face book, LinkedIn and twitter platform. The research performed by them was finding profile of single person on multiple platforms. Idea about to study this paper is finding the way of matching profile, its parameter and concern about matching. They have worked with personal identity, relation identity and social identities to find its id on another platform also. They have also taken location, name and user consideration connected with them. Further similarity rank finding and matching is performed which help in finding matching identities in multiple platforms. This concept can similarly draw in single platform to find its multiple identities.

In this paper [10] author proposed a graph based theory for Sybil detection and prevention. Graph based theory which is based on vertex and edges is proposed which help in calculating distance among the entity and provide their score value to be a Sybil. They have considered undirected, unweighted, non-bipartite and all the connected graph for Sybil detection. Detection model proposed by them give the detection value and labeled identity matching over the given inputs. Finally number of attack edges and positive ratio calculation is performed in their research.

SybilShield is another OSN-based protocol that assumes that a network consists of a number of small, medium, and large communities. SybilLimit and SybilGuard, on the other hand, assume that a network consists of a respective legitimate and Sybil region. SybilShield's other assumption is that two given networks are fast mixing, and a malicious attacker can create numerous identities, but few trusted relations exist between a legitimate node and a Sybil node. The protocol defines an edge between different communities as a foreign edge, meaning that edges formed between legitimate and illegitimate communities are fewer than the number of edges formed between legitimate communities.

Integro is an extension of SybilRank and likewise studies the posted content. The authors of [69] proposed Integro as a better solution for detection of Sybils. They stated that this method achieved 95% precision in Sybil detection, whereas SybilRank achieved 43%. The method was tested on Facebook, RenRen, and Tuenty, and it proved to be effective. Integro uses various node features to detect Sybils and identify potential victims in a non-adversarial setting. The developers of the method employed only known legitimate profiles, which accepted or rejected friendship requests sent by known Sybil profiles. The authors did not accept the assumption that Sybil nodes can have a limited quantity of friends and therefore attack edges. The authors first proposed detection of victim accounts based on user-level activities. They acknowledged that many users are not cautious in OSNs and accept friend requests from users whom they do not know, especially if they have common friends. They model social networks as an undirected graph with the nodes representing user accounts and the edges representing a bilateral social relationship among nodes. The nodes have a degree equal to the sum of the weights on their edges. The set of nodes is divided into two parts: real and fake accounts. Then, a Sybil region is detected. Integro sets a feature vector for each node to predict the probability that the user will be a victim. It also counts a vulnerability score, which represents the level of the probability that a user will become a victim. Using the power iteration method, Integro computes trust values.

Problem Formulation

In existing technique, there are few limitations which recommended solving for proper identity matching and prevention:

1. A false detection rate is observed, which often give wrong effect from the detection phase. Thus a high accuracy and proper detection rate is required to study.
2. Limited parameter consideration keeps limited detection and roles which further need to be investigate and multiple parameters need to apply.
3. Previous approach worked with the data share, data utilized by the identity, further a spatial data can be involved to detect proper identity from the list of identities.
4. A high usage and selfish identity detection is need to investigated.

Proposed Work

In order to work further with the synonym identity detection and prevention an approach with spatial data analysis, multiple parameter weight and score computation. Using this weight parameter in Sybil detection is going to proposed in further detection and role identity mechanism.

Conclusion

Web social platform and their gaining popularity is gradually increasing. The entire platform also providing advertising platform for multiple organizations. Social network help in boosting advertisement with paid and organic manner. Some entity make use of these platform and try to put multiple identity of same business and thus an over impression, false result get generate for the user. This paper survey previous mechanism which participates in data sharing Sybil identity detection and algorithm worked towards the duplicate identity detection and prevention approach. Our further investigation is going to find a proposed enhance mechanism which can provide better accuracy than past algorithms.

References

[1] Teerapol Silawan and Chaodit Aswakul," SybilVote: Formulas to Quantify the Success Probability of Sybil Attack in Online Social Network Voting", DOI 10.1109/LCOMM.2017.2687867, IEEE.
[2] P. Anagha and J. Krishnan, "Vote credence: Social network sybil defence by user behaviour," International Journal of Science and Research (IJSR), vol. 5, no. 6, pp. 1500–1503, 2016.
[3] Y. Hu, J. Song, Min Chen, "Modeling for Information Diffusion in Online Social Networks via Hydrodynamics", IEEE Access, accepted, 2016.
[4] M. Saud Khan and N. M. Khan, "Low complexity signed response based sybil attack detection mechanism in wireless sensor networks," Journal of Sensors, vol. 2016, 2016.
[5] H. Shen and X. Liu, "Detecting Spammers on Twitter Based on Content and Social Interaction," in Network and Information Systems for Computers (ICNISC), 2015 International Conference on, 2015, pp. 413- 417.
[6] K.M.Ponsurya, R.Poornima, Mrs.S.Vairachilai, "Transparent user identity and overcoming Sybil attack for secure social networks", 2015 International Conference on Computer Communication and Informatics (ICCCI -2015), Jan. 08 – 10, 2015.
[7] N. Abokhodair, D. Yoo, and D. W. McDonald, "Dissecting a Social Botnet: Growth, Content and Influence in Twitter," in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015, pp. 839-851.
[8] Reza Soltani, Abdolreza Abhari, "Identity Matching In Social Media Platforms", IEEE, 2013.
[9] Y. Boshmaf, "A Quick Survey of Social Network-based Sybil Defenses," 2013.
[10] Yazan Boshmaf, Konstantin Beznosov, Matei Ripeanu," Graph-based Sybil Detection in Social and Information Systems", 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
[11] D. Correa, A. Sureka, and R. Sethi, "WhACKY!-What anyone could know about you from Twitter," in Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference on, 2012, pp. 43-50.
[12] H. Saif, Y. He, and H. Alani, "Alleviating data sparsity for twitter sentiment analysis," in The 2nd Workshop on Making Sense of Microposts, 2012.
[13] L. Guo, Y. Fang, and L. Wei, "Fine-grained privacy-preserving reputation system for online social networks," in 2013 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2013, pp. 230–235.