# DETECTION OF SYBIL USERS USING RANDOM WALK THEORY

**[1]PRIYA CHANDAPILLAI, [2]SUDHIR D. SAWARKAR**

[1,2]Datta Meghe College of Engineering, Datta Meghe College of Engineering,
Mumbai University, Mumbai University
E-mail: [1]chandapillaipriya@gmail.com, [2]sudhir_sawarkar@yahoo.com

**Abstract**— Large-scale decentralized systems without trusted identities and other peer-to-peer systems are more likely exposed to Sybil attacks from remote faulty elements that compromise the system's running with fake information. In such a Sybil attack, a malicious user pollutes the system by creating multiple fake identities, pretending to be multiple, distinct nodes in the system called Sybil nodes. Many security mechanisms against such attacks proposed earlier are based on specific assumptions. In this paper, a novel defense mechanism against the sybil attacks on social network is proposed using random walk beginning from a known honest node traversing throughout the social graph of the network. Besides we propose social degree and popularity analysis on the detected honest users obtained from the random walk approach which will eventually give the list of Sybil users and an action can be taken to block such Sybil users. This paper aims for improved Sybil detection in the social network as it mitigates attacks from friends of friends, thereby reducing the redundancy of data and fake identities in the network and also outperforms the running of the network.

**Index Terms**— Sybil Attacks, Random Walk, Social Degree, Popularity Analysis.

## I. INTRODUCTION

Large-scale peer-to-peer systems or distributed systems allow cooperating users to enjoy a service without the need for any centralised infrastructure unlike conventional client-server systems. P2P systems have no central server and it becomes extremely easy for such systems to extend and gets vulnerable to various attacks. These distributed systems are exposed to threats from faulty or hostile remote computing elements.

Social networks are also peer-to-peer systems with popular infrastructures for communication, socializing, and information sharing on the Internet. Users form social links to friends, and leverage these links to share information, organize events, and search for specific users or shared resources. Digg, Youtube, Facebook and BitTorrent are open systems that allow any user on the Internet to join the system easily and communicate through emails and quick messaging and collaborate through content rating and recommendation. Therefore a need of strong user identity in the distributed systems is required to remove the vulnerability against sybil attacks.

In a sybil attack, a malicious user pollutes the system by obtaining multiple fake identities and pretends to be multiple, distinct nodes in the system. Malicious users can compromise the running of the network by generating and controlling large numbers of shadow identities. This attack is named after a patient 'Sybil Dorsett' suffering from dissociative identity disorder who manifested 16 different personalities. Practical examples of such attacks are Rig Internet polling by using multiple IP addresses for submitting votes and to increase Google Page-Rank rating of a page, sharing of iTunes passwords for shared media access and On-line forums, starting with USENET to contemporary blogs or virtual worlds like Second Life always have to deal with the problem of disruption in the discussion threads, with persistent abusers coming back under different names.

Sybil attacks can be mitigated by using centralised defence mechanism i.e. assuming the existence of a trusted authority which could rate-limit the introduction of fake identities by asking the users to verify with some credentials, like social security number, or by making some payment. However, such requirements will prevent users from accepting such systems, as they impose additional burdens on users. Decentralised mechanism i.e. without a trusted central authority is much harder as many systems today try to combat sybil attacks by binding an identity to an IP address. However, malicious users can readily harvest or steal IP addresses, spammers, for example, are known to harvest a wide variety of IP addresses to hide the source of their messages.

In a social network, two user identities form a link after establishing a relationship between them. In the social graph each identity is represented as a node. To prevent the adversary from creating many sybil identities, the previous defence mechanisms are built upon the assumption that the number of links between the honest nodes and the sybil nodes, i.e. the *attack edges*, is limited. As a result, although an adversary or fake identity can create many sybil nodes and link them in an arbitrary way, a *small cut* will be formed between the honest region and the sybil region. The small cut includes all the attack edges and if removed will disconnect the sybil nodes from the rest of the graph.

All social network-based Sybil defence schemes make the assumption that an attacker may create arbitrary sybil identities in social networks but no one can establish an arbitrarily large number of social

connections to non-Sybil nodes.

## II. LITERATURE SURVEY

The sybil attack [1] is a powerful threat faced by any decentralized p2p system that has no central, trusted authority that will vouch for a one-to-one correspondence between users and identities. This research already proved a series of negative results. One promising way is to leverage the social network topologies to defend against sybil attacks in social networks. SybilGuard [2] and SybilLimit [3], which are two decentralised algorithms, both rely on the assumption that social networks are fast mixing and the number of attack edges is limited. To identify sybil nodes, these defence schemes make use of random routes, in which each node uses a pre-computed random permutation as a one-to-one mapping from incoming edges to outgoing edges. SybilGuard suffers from high false negatives, as each attack edge may introduce O(√n log n) sybil nodes without being detected and SybilLimit, reduces this value to O(log n). Moreover, to detect the sybil region with SybilGuard or SybilLimit, all the suspect nodes in the social graph need to be tested. GateKeeper [5] is another decentralized sybil defense scheme that heavily relies on the assumption that the social networks are random expander and suffers from high false positive and negative rates and cannot effectively identify sybil nodes on the real-world asymmetric social topologies. SybilInfer [4], a centralized sybil defense algorithm, leverages a Bayesian inference approach that assigns a Sybil probability, indicating the degree of certainty, to each node in the network. It achieves low false negatives at the cost of high computation overhead. The overall time complexity of SybilInfer is O(|V |2 log |V |), where V is the set of vertices in the social graph. In the evaluation SybilInfer handled networks with up to 30K nodes, which is much smaller than the size of regular online social networks. SybilDefender [6], a centralised Sybil defence mechanism consists of a sybil identification algorithm and a sybil community detection algorithm. It outperforms SybilLimit, by one to two orders of magnitude in both accuracy and running time. SybilDefender though relies on performing a limited number of random walks requires more of processing time. [12]Detects attack edges, and then prohibits the communication over the detected edges. SRNC, has two stages, Aggregation and ReAggregation where the weight of each edge is computed and the suspect edges – the edges with high weights, are identified. It ensures honest peers accept O(|AE|) sybil peers i.e. O(log(n)) times improvement over SybilLimit where n is the number of peers and |AE| is the number of attack edges. We propose two approaches in Sybil detection with reduced processing time and filteration on even detected sybil and non-sybil nodes.

## III. PROPOSED SYSTEM

Our proposed work is designed to defend the sybil attacks and to detect the sybil users in the social network using random walk theory. Further, based on social degree and popularity analysis the system tries to improve the detection of sybil users in the network enabling more specific output.

The social network is considered as a graph G (V,E), where the nodes in the vertices V are the honest or sybil users while the relationship between two corresponding nodes is connected using an edge E. The honest users have one single identity while the malicious users have multiple Sybil identities.

Our Proposed system has two approaches to detect the sybil users in the social network: a random walk approach, and a social degree and popularity analysis. To mitigate sybil attacks using random walk is also used in several defence schemes. But to get improved detection of sybil users along with random walk a mechanism has to be devised which will detect malicious users even among friends of friends in a social network. This is because the malicious behaviour can be present amongst any of the user, so we need to analyze a group of users i.e. friends of friends. This imbibes a need to design a system to defend sybil attacks with at most efficiency and performance. The two approaches together will help achieve this.

### *A. Random Walk Approach*

A random walk on a social graph G (V, E) is defined by the sequence of moves of a particle between nodes of G. In a random walk, at each hop, the current node selects a uniformly random edge to direct the walk. Our system has to traverse the network using random walk from a known honest node *h* and find all the lengths of the paths of random walk, get the maximum and minimum lengths as $l_{max}$ and $l_{min}$. We define the frequency of a node as the number of times the node is being traversed by random walk. Get *m* as the count of nodes whose frequency is compared to a threshold *t*. The algorithm computes the mean and standard deviation for each length and outputs *<l, mean, stdDeviation>*. Then the algorithm compares *m* and *mean* by a value stdDeviation * $\alpha$ *(*where $\alpha$ is the level of accuracy required in deviation) and outputs whether a node is sybil or not.

### *B. Social Degree and Popularity Analysis*

The detected honest users from random walk approach act as an input to this approach where parameters like social degree and popularity is computed. The social degree is the total number of friends of a user and popularity is the number of visits the user has. It is essential to analyze a group of users so the social degree of friends of friends is computed. Thereafter the average of social degree and popularity is calculated

for the network. These parameters are then compared with the friend count of each user and the profile visits and activity. This will give a list of sybil users. To improve the efficiency again a filtration is done on the sybil and non-sybil users obtained. For which, an intersection count is computed to check if there are any common nodes from the random walk approach. If the count is less than 50% of the sybil node, then such nodes will be removed from the sybil list and if the intersection count is more than 50% of the non-sybil nodes, then such nodes has to be added into the sybil list. The admin shall label the users as honest or sybil and accordingly the decision is taken to block the detected sybil users.

## IV. SYSTEM ARCHITECTURE

In proposed system, a user is created in the system by taking into consideration certain essential data field and is authenticated by the system by comparing and validating his log in credentials. When a person is authenticated perfectly by the system, he can add other users as his friend in the system. Before adding a friend, the system sends a request to the particular friend for his approval. Once the friend approves the request, user reflects as a friend in the current user's screen. The user can also view and add posts and also store the tracking parameters i.e. social degree and popularity.

The admin has the right to view all users in the system and manages all the user data. For the sybil detection, the application calculates the parameters like length, mean and standard deviation from random walk. Later the parameters for social degree and popularity analysis are calculated and all the sybil users are detected.

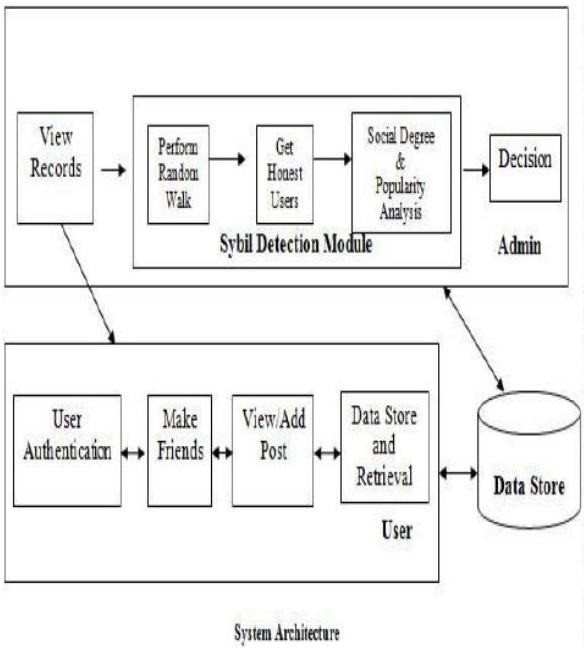


**Fig. 1: System Architecture**

## V. ALGORITHM

1: Consider *h* as honest node and *S* as suspect node.
2: Find *n* i.e. total no of nodes or users in the system then select *f* (short random walks) start from honest node and up to *particular ending node in current path*
3: **for** all nodes as *i*
4: Calculate $Freq_i$ for all paths
5: **end for**
6: Threshold *t* = 25% * (sum ($Freq_i$) / *n)*
7: Take all random walks in one array (*j*) and find the shortest one as ($l_{min}$) and largest one as ($l_{max}$) path from that array.
8: **while** $l <= l_{max}$
9: Identify paths of length *l* in current graph
10: **for all** nodes as *i*
11: Calculate $LengthFreq_i$ for all paths of length *l*
12: ***end for***
13: Calculate *mean* and *std deviation* of frequencies calculated for all nodes for length *l*
14: Then we will get output ($l$, $mean_l$, $stdDeviation_l$)
15: ***end while***
16: **while** *l* <= *lmax*
17: Calculate no of nodes *m* greater than *t*
18: If *mean*− *m* > *stdDeviation*(*l*) * $\alpha$ then we can say that, length *l* is sybil at end of loop otherwise *l*++ and it will be honest.
19: Where $\alpha$ is level of accuracy required in deviation (default=0.3)
20: **end loop**
21: if *u* is honest perform s*ocialDegree* and *Popularity analysis*
22: *Popularity* = sum (no. of profile visits by friend i) i from 0 to no. of friends
23: *socialDegree (x)* = | *F(x)* |
24: *U*: the set of all users
25: *Popularity(i)* : the popularity of the user *i* ($i \in U$)
26: *socialDegree(i)* : the social degree of the user *i* ($i \in U$)
27: *Ci* is No. of friends of neighbor *j* (second level friend count)
28: *Fi*: the set of user *i*'s friends ($i \in U$)
29: Calculate average of *Popularity* and *Social Degree*.
30: for ($i \in U$) {
31: **if** (*Popularityi* <= *avg_pop*) {
32: **if** (*socialDegreei(Ci*) >= *avg_socialDegree*)
33: {
34: $S = S \cup \{i\}$
35: }
36: }
37: }
38: **for** ($i \in S$)
39: {

```
if (| Fi ∩ S | < 0.5 * |S|)
S = S − {i}
}
}
for (i ϵ U)
{
if (|Fi ∩ S | > 0.5 * | Fi |)
{
S = S ∪ {i}
}
}
return S;
```

## CONCLUSION

In this paper, a new defense mechanism against sybil attacks in social networks is proposed. Out of the two approaches discussed here to detect sybil users in the network, one is to use the random walk in the social graph beginning from the very first known honest node. The second approach is to analyze the detected honest users and perform social degree and popularity computations on those users which will eventually give the list of sybil users and an action can be taken to block such sybil users. The proposed work ensures improved sybil detection in the social network thereby reducing the redundancy of data and fake identities in the network. Our proposed work will be scalable to large networks. It shall also improve the efficiency and shall outperform the running of the network.

## REFERENCES


[1] J. R. Douceur. The sybil attack. In *IPTPS*, 2002.

[2] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. SybilGuard: defending against sybil attacks via social networks. In *SIGCOMM*, 2006.

[3] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. SybiLimit: A near optimal social network defense against sybil attacks. In *IEEE Symposium* on Security and Privacy, 2008.

[4] G. Danezis and P. Mit. SybilInfer: Detecting sybil nodes using social networks. In NDSS, 2009.

[5] N. Tran, J. Li, L. Subramanian, and S. S.M. Chow. Optimal sybilresilient node admission control. In IEEE INFOCOM, 2011.

[6] Wei Wei∗, Fengyuan Xu∗, Chiu C. Tan†, Qun Li∗. SybilDefender: Defend Against Sybil Attacks in Large Social Networks

[7] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *ACM/USENIX IMC*, 2007.

[8] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove. An analysis of social network-based sybil defenses. In *SIGCOMM*, 2010.

[9] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *EuroSys*, 2009.

[10] K. Xing and X. Cheng. From time domain to space domain: Detecting replica attacks in mobile ad hoc networks. In *IEEE INFOCOM*, 2010.

[11] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in facebook. In *SIGCOMM WOSN*, 2009.

[12] L. Xu, S. Chainan, H. Takizawa, and H. Kobayashi. Resisting Sybil attack by social network and network clustering. In *SAINT*, 2010.

[13] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *WWW*, 2009.

[14] M. Mitzenmacher and E. Upfal. *Probability and Computing*. Cambridge University Press, 2005.

[15] Rate your relationships. http://apps.facebook.com/ratingrelationships/.

[16] Amazon mechanical turk. https://www.mturk.com/mturk.

[17] The top 500 sites on the web. http://www.alexa.com/topsites.

[18] R. Albert and A. Barab´asi. Statistical mechanics of complex networks.*Rev. Mod. Phys*, 74:47–97, 2002.

[19] P. Erd¨os and A. R´enyi. On random graphs. *Publicationes Mathemticae (Debrecen)*, 6:290–297, 1959.

[20] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," ACM Trans. Prog. Languages Syst., vol. 4, no. 3, pp. 382–401, 1982.

[21] A. Ramachandran and N. Feamster, "Understanding the network-level behavior of spammers," in Proc. ACM SIGCOMM, 2006, pp. 291–302.

[22] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: Membership, growth, and evolution," in Proc. ACM KDD, 2006, pp. 44–54.

[23] S. Wasserman and K. Faust, Social Network Analysis. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[24] R. Bazzi and G. Konjevod, "On the establishment of distinct identities in overlay networks," in Proc. ACM PODC, 2005, pp. 312–320.

[25] A. Cheng and E. Friedman, "Sybilproof reputation mechanisms," in Proc. 3rd ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems (P2PECON-05), Philadelphia, PA, Aug. 2005, pp. 128–132.

[26] M. Richardson, R. Agrawal, and P. Domingos, "Trust management for the semantic web," in Proc. 2nd Int. Semantic Web Conf. (ISWC2003), Sanibel Island, FL, Oct. 2003, pp. 351–368.


★★★