DETECTION OF SYBIL NODES IN A SOCIAL NETWORK USING ANTI-SYBIL MECHANISM

G. Lawrence Paul Sundararaj[1] N. Subbulakshmi[2]

[1]PG Scholar, [2]Assistant Professor, Department of IT,

Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, (India)

ABSTRACT

The majority of the substantial scale long range interpersonal communication locales and little private informal communities on the Internet are interested in Sybil assaults. The Sybil assault is an assault where in a foe makes various Duplicate or False characters to bargain the running of the framework. By including false data by the Duplicated elements, an enemy can delude a framework into settling on choices profiting. Guarding against Sybil assaults is very difficult. This paper presents Sybil Defender, a sybil safeguard instrument that influences the system topologies to shield against sybil assaults in interpersonal organizations. Sybil Defender can viably distinguish the sybil nodes and identify the sybil group around a sybil node and it is practical to breaking point the quantity of assault edges in online interpersonal organizations by relationship rating.

***Keywords*—Sybil Attack, Social Network, Random Walk**

I. INTRODUCTION

Most systems, in the same way as a shared system, depend on suppositions of personality, where every PC speaks to one character. A Sybil assault happens when a frail node is seized to claim different characters, as such, a Sybil attack[1] happens when the aggressor makes numerous personalities (sybils) and abuses them so as to control a notoriety score. Issues arise[2] when a notoriety framework, (for example, a record imparting notoriety on a torrent system) is deceived into feeling that an assaulting node has a disproportionally substantial impact. So also, an assailant with numerous characters can utilize them to act vindictively, by either taking data or upsetting correspondence. It is imperative to perceive a Sybil assault and note its threat so as to shield yourself from being a target.

Recently, there has been an expanding enthusiasm for protecting against sybil assaults in informal communities [3], [4], [5], [6], [7]. In an informal organization, two client personalities impart a connection if a relationship is made between them. Every character is spoken to as a node in the social chart. To keep the enemy from making numerous sybil characters, all the past sybil guard plans are based upon the supposition that the quantity of connections between the sybil nodes and the fair nodes, otherwise called assault edges, is constrained. Therefore, in spite of the fact that an enemy can make numerous sybil nodes and connection them in a self-assertive path, there will be a little cut between the legitimate locale and the sybil district. The little cut comprises of all the assault edges and its expulsion detaches the sybil nodes from whatever remains of the diagram, which is utilized by past plans to distinguish the sybil nodes. Note that the answer for this issue is nontrivial, on the grounds that discovering little cuts in a chart is a NP-hard issue. To point of confinement the

quantity of assault edges, past plans expect that all the connections in interpersonal organizations are trusted and they mirror the trust connections among those clients in this present reality, and subsequently, a foe can't create numerous associations with the legit clients. Be that as it may, it has been demonstrated that this presumption does not hold in some genuine interpersonal organizations [8].

To address the issue Sybil Defender is presented, which is an incorporated sybil safeguard system. It comprises of a sybil distinguishing proof calculation to recognize sybil nodes, a sybil group location calculation to distinguish the sybil group encompassing a sybil node, and two ways to constraining the quantity of assault edges in online interpersonal organizations. Our plan is in view of the perception that a sybil node must experience a little slice in the social diagram to achieve the fair area. A genuine node, in actuality, is not limited. Presently, in the event that we begin from a sybil node to do irregular strolls, the arbitrary strolls have a tendency to stay inside the sybil district. The primary commitments of this work include:

i. Based on performing a predetermined number of arbitrary strolls inside the social charts, our proposed sybil ID and sybil group identification algorithms are more proficient than past procedures for expansive interpersonal organizations.
ii. We assess SybilDefender for tests. The outcomes demonstrate that the execution of our sybil ID calculation approaches the hypothetical bound, and it outflanks SybilLimit, the best in class sybil safeguard system that applies to extensive informal organizations, by more than 10 times in both exactness and running time. Furthermore, our sybil group location calculation can viably distinguish the sybil group around a sybil node with short running time.

We propose two useful systems to point of confinement the quantity of assault edges in online interpersonal organizations, and add to a Facebook application to exhibit the achievability of one of the methods. The overview aftereffects of our Facebook application demonstrate that the presumption made by past work that all the connections in informal communities are trusted does not hold in online interpersonal organizations, and it is doable to utmost the quantity of assault edges in online informal organizations by relationship rating.

II. RELATED WORK

Sybil Guard [7] and Sybil Limit [6] are among the first Sybil identification plans to be proposed. Sybil Guard utilizes the convergences between modified arbitrary strolls to figure out if characters ought to be offered access to the framework. Sybil Limit enhances Sybil Guard's bound by utilizing numerous strolls, which permits it to acknowledge less Sybil personalities every assault edge. Both of these plans can be executed in a concentrated or decentralized style.

Sybil Infer[3] is an incorporated convention that expect full learning of the social chart. It utilizes a Bayesian deduction procedure that allots to every node its likelihood of being Sybil. Not at all like Sybil Guard and Sybil Limit, does Sybil Infer not give any hypothetical limits on the quantity of Sybil characters acknowledged every assault edge. In the assessment, Sybil Infer took care of systems with up to 30,000 nodes, which is much littler than the measure of standard online interpersonal organizations.

Gatekeeper [23] is a decentralized Sybil discovery convention that enhances over the insurances gave by Sybil Limit. It utilizes a variation of the ticket dispersion calculation utilized as a par with it.

Sum Up [22] from different irregular personalities in the chart to recognize Sybils. Despite the fact that informal organization based Sybil location plans are moderately straightforward and simple to coordinate into the framework, they all experience the ill effects of innate limits. Specifically, these plans make solid

presumptions about the topology of the social diagram, where a large portion of certifiable informal communities don't comply with these suppositions. Thus, these plans have not discovered standard adaption, and they normally bring about high false positive and false negative rates in certifiable interpersonal organizations. Conversely, Sybil Defender just depends on performing a set number of irregular strolls in the social chart, and it is adaptable to huge systems.

III. PROPOSED FRAMEWORK

We mean the informal community as a chart G comprising of vertices V and edges E. There are n fair clients in the informal organization, each with one personality, meant as a legit node in V. There are likewise one or more vindictive clients in the informal community, each with various Sybil personalities. Every Sybil personality is indicated as a Sybil node in V. A relationship between two personalities in the informal community is spoken to as an edge interfacing the two comparing nodes in G. The edges in G are undirected. We name the edge between a Sybil node and a genuine node an assault edge. The Sybil area comprises of all the Sybil nodes, while the legitimate district comprises of all the genuine nodes. All the Sybil nodes are controlled by a foe. Accordingly, the enemy can make self-assertive edges inside the Sybil locale.

Graphs[10] are utilized to speak to relation siphs or associations (edges) between area objects (vertices). Informal communities are a use of a diagram information structure.

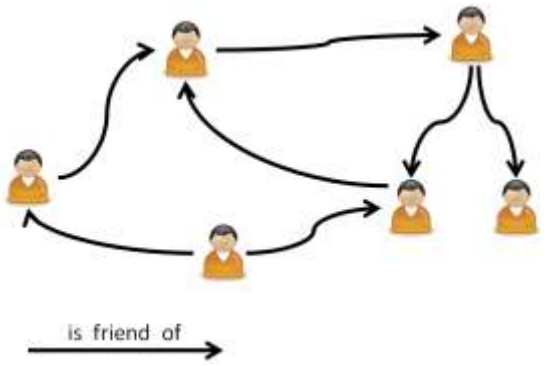

Figure 1 A Social Network

The proposed approach, Sybil Defender is a popular methodology for hostile to sybil. It comprises of two steps: Sybil node discovery and Sybil Community detection (Fig 2). One approach to shield against sybil assaults in informal organizations is to influence the interpersonal organization topologies. This paper is expand on the accompanying suppositions:

The fair district is quick blending, which implies an irregular stroll of length O (log n)is sufficiently long such that with likelihood no less than 1-1/n, the last crossed node is drawn from the node stationary appropriation of the chart.

1. One honest node is known.
2. The social network topology is known
3. The size of the sybil region is not comparable to the size of the honest region.
4. the number of attack edges is limited.

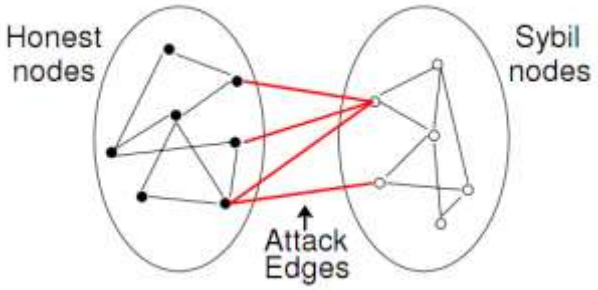

FIGURE 2 An Honest Community and Sybil Community

In a social network, the vertices (nodes) are identities in the distributed system and the (undirected) edges correspond to human-established trust relations in the real world. The edges connecting the honest region (i.e., the region containing all the honest nodes) and the Sybil region (i.e., the region containing all the Sybil identities created by malicious users) are called attack edges.

3.1 Sybil Node Detection

In Sybil node recognition, the Sybil recognizable proof calculation that takes the social diagram G (V, E), a known legit node h, and a suspect node u as information, and yields whether u is Sybil or not. Our calculation is in view of irregular strolls. An irregular stroll on a diagram is characterized by the succession of moves of a molecule between nodes of G.

For a suspect node, in view of preknown legit nodes' factual peculiarities, Sybil Defender figures out if the suspect node is a Sybil or not. In the wake of discovering a Sybil node, in view of the suspicion that Sybil nodes are more prone to associate with other Sybil nodes, the guard will recognize the Sybil group in which the Sybil node resided.

This defense is based on two assumptions:

(1) the number of links between honest users and Sybils is limited and

(2) the size of the Sybil community is smaller than that of millions of users: for the attacker, to register such a large number of identities is impossible.

From an honest user, we can send a fixed number of random walkers to pass an *l*-length random path, assuming there are k walkers. At other nodes, we can compute the times that these random walkers passed through this node and call the times their *visiting frequency*. After that, we can calculate the statistical distribution of the visiting frequency. If the random walks from a suspect node do not follow some statistic distribution, then the suspect is Sybil.

3.2 Sybil Community Detection

After one Sybil node is identified, our Sybil community detection algorithm can be used to detect the Sybil community surrounding it. The Sybil community detection algorithm takes the social graph *G (V, E)* and a known Sybil node s as input, and outputs the Sybil community around s. Our algorithm relies on performing partial random walks originating from s. Each partial random walk behaves the same as the simple random walks used in the previous section, except that it does not traverse the same node more than once. Therefore, when a partial random walk reaches a node with all the neighbors traversed by itself, this partial random walk is "dead" and cannot proceed. Figure 3 illustrates the Sybil Defender. Suppose that we have already known an honest node. From this node, we send out *k* random walks with a fixed length *l*. Since social network (in the honest region) is fast mixing, which means that any pair of nodes can reach one another at an *O*(log n)-length

random path, a circle region in the honest community will be covered by the random walk. However, because the size of the Sybil community is smaller than that of honest one, the majority of random walks in the Sybil region is different from the honest one. By this way, a suspect node can be verified.

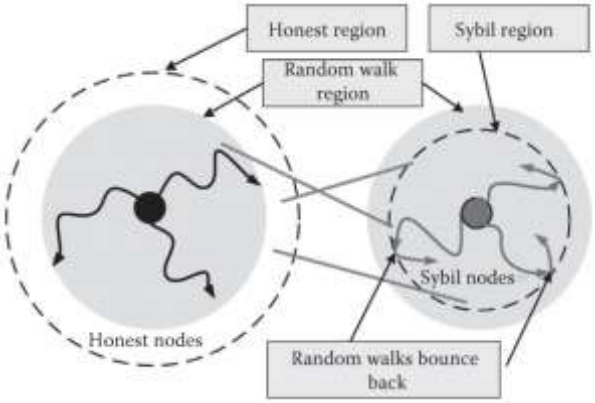

FIGURE 3 The Idea of Sybil defender.

After finding a Sybil node, the Sybil Defender can also detect its resident Sybil community, based on the fact that Sybil nodes are more likely to connect with other Sybil nodes. The detection of a Sybil community can be done by using loop-free random walks. Consider that when a random walker passes the same node twice, it means the random walkers reach the boundary of the Sybil community. Sybil Defender renders a random walker dead if it arrives at the same node twice. Similar to the process of verifying a suspect node, Sybil Defender also initializes several random walks with different lengths. Again, the reason is that the size of the Sybil community is unavailable. If the dead ratio of the *L*-length group of random walks is greater than a predefined threshold, then all the passed nodes will be regarded as members of a Sybil community.

Algorithm 1 is used to calculate the mean number of nodes with frequency no smaller than t when performing R random walks in length of l originating from known honest node h.

The larger the community is, the larger mean number it gets.

Algorithm 1. PreProcessing(G, h).

1: $J = \{h\}$
2: **for** $i = 1$ to f **do**
3: Perform a random walk with length $l_s = \log n$ originating from h
4: $J = J \cup \{the\ ending\ node\ of\ the\ random\ walk\}$
5: **end for**
6: $l = l_{min}$
7: **while** $l <= l_{max}$ **do**
8: **for** $i = J.first()$ to $J.last()$ **do**
9: Perform R random walks with length l originating from node i
10: Get n_i as the number of nodes with frequency no smaller than t
11: **end for**
12: output $\langle l, mean(\{n_i : i \in J\}), stdDeviation(\{n_i : i \in J\})\rangle$
13: $l = l + 100$
14: **end while**

Algorithm 2 is to detect whether a given node is a sybil node or not. As is said in above, if a node has a very low number, it may be a sybil node.

Algorithm 2. SybilIdentification($G, u, tuples\ from\ Alg.1$).

```
1: l = l_0
2: while l <= l_max do
3:     Perform R random walks with length l originating
       from u
4:     m = the number of nodes whose frequency is no
       smaller than t
5:     Let the tuple corresponding to length l in the outputs
       of Algorithm 1 be ⟨l, mean, stdDeviation⟩
6:     if mean − m > stdDeviation ∗ α then
7:         output u is sybil
8:         end the algorithm
9:     end if
10:    l = l ∗ 2
11: end while
12: output u is honest
```

Partial random walk is a little different from regular random walk: once a node is 'walked', it cannot be walked once again. So a partial random walk can be terminated when a node have no neighbors to walk into without reaching a given length. It is called a 'dead walk'. Given a sybil node s, we can estimatioin a length when the dead Walk Ratio is smaller than a given threshold. All the nodes within a length less than the estimated length are suspected to be sybil nodes.

Algorithm 3. WalkLengthEstimation(G, s).

```
1: l = l_0/2
2: deadWalkRatio = 0
3: while deadWalkRatio < β do
4:     l = l ∗ 2
5:     deadWalkNum = 0
6:     for i = 1 to R do
7:         Perform a partial random walk originating from s
           with length l
8:         if the partial random walk is dead before it
           reaches l hops then
9:             deadWalkNum++
10:        end if
11:    end for
12:    deadWalkRatio = deadWalkNum/R
13: end while
14: output l
```

Rate all the suspected node based on conductance, defined as follows. Let d be the sum of the degrees of all the nodes in set S, and a be the number of edges with one endpoint in S and one endpoint in S'. The the conductance of S is a/d. Based on the assumptions, there should be a small cut between the honest region and sybil region. The conductance of the sybil region is very small. So it is reasonable to leverage a greedy algorithm to detect which nodes are real sybil nodes.

Algorithm 4. SybilRegionDetection($G, s, l\ from\ Alg.3$).

```
1: Set the frequency of all the nodes to be 0
2: for i = 1 to R do
3:     Perform a partial random walk originating from
       node s with length l
4:     s.frequency++
5:     for j = 1 to l do
6:         Let the j^th hop of the partial random walk be
           node k
7:             k.frequency++
8:     end for
9: end for
10: traversedList = Sort the traversed nodes by their
    frequency in decreasing order
11: counter = 0
12: S = ∅
13: do
14:     counter = conductance(S)
15:     for i = traversedList.first() to traversedList.last()
        do
16:         if node i ∈ S then
17:             continue
18:         if conductance({i} ∪ S) <= conductance(S) then
19:             S = {i} ∪ S
20: while (counter > conductance(S))
21: output S
```

The Sybil node have a tendency to be before the legitimate hubs in the sorted rundown, in light of the fact that an expansive number of fractional irregular strolls can't enter the legit locale, because of the presence of the little cut between the fair district and the Sybil area. Thus, the covetous calculation will first attempt to include the hubs that are more probable Sybil to S. This calculation just depends on performing R halfway arbitrary strolls starting from a Sybil hub, which makes it extremely effective and versatile to extensive estimated informal organizations.

IV EXPERIMENTAL RESULTS

We ponder the conduct of Sybil Defender when there are malevolent clients. In most security research, the expression "pernicious client" commonly alludes to a solitary malignant client who does not expect extra personalities. In arbitrary, we over and over pick consistently irregular hubs in the diagram as Sybil assailants, until the aggregate number of assault edges achieves a certain worth. The exploratory after effects of our proposed approach demonstrate the precision and the effectiveness level a bit higher than the current ones. Additionally the Sybil hub location rate and time utilization likewise has a tendency to be high. False distinguishing proof rate are less of likelihood with our Sybil Defender. We have gotten all relating results for Sybil group also, which are dependably marginally better yet the distinction is typically insignificant.

V. CONCLUSION

The test with the Sybil Defender is that how to concentrate the right going to recurrence dispersion from the legitimate locale. This paper introduced Sybil Defender, a concentrated Sybil safeguard system against sybil assaults utilizing informal organizations. Sybil Defender comprises of a sybil ID calculation, a sybil group recognition calculation, and two ways to restricting the quantity of assault edges in online informal communities. Our outcomes on informal communities confirmed their quick blending property, and therefore approved the principal suspicion behind this methodology. As future work, we plan to actualize Sybil Limit inside the connection of some true applications and show its utility. As future work, we intend to implement Sybil Limit within the context of some real-world applications and demonstrate its utility.

ACKNOWLEDGEMET

I extend my sincere thanks to D. R. Anita Sofia Liz for their innovative ideas and suggessions throughout my project and the successful completion of my project.

REFERENCES

[1] http://itlaw.wikia.com/wiki/Sybil_attack.
[2] http://anti-virus-software-review.toptenreviews.com/what-is-a-sybil-attack-.html.
[3] G. Danezis and P. Mit, “Sybilinfer: Detecting Sybil Nodes Using Social Networks,” Proc. Network and Distributed System Security Symp. (NDSS), 2009.
[4] N. Tran, J. Li, L. Subramanian, and S.S. Chow, “Optimal Sybil-Resilient Node Admission Control,” Proc. IEEE INFOCOM, 2011.
[5] L. Xu, S. Chainan, H. Takizawa, and H. Kobayashi, “Resisting Sybil Attack by Social Network and Network Clustering,” Proc. IEEE/IPSJ 10th Int’l Symp. Applications and Internet (SAINT), 2010.

[6] H. Yu, P.B. Gibbons, M. Kaminsky, and F. Xiao, "SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks," Proc. IEEE Symp. Security and Privacy, 2008.

[7] H. Yu, M. Kaminsky, P.B. Gibbons, and A. Flaxman, "SybilGuard: Defending against Sybil Attacks via Social Networks," Proc. ACM SIGCOMM, 2006.

[8] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks," Proc. 18th Int'l Conf. World Wide Web (WWW '09), 2009.

[9] N. Tran, B. Min, J. Li, and L. Subramanian. Sybil-resilient online content voting. In Proceedings of the 6th USENIX symposium on Networked systems design and implementation, NSDI'09, pages 15– 28, Berkeley, CA, USA, 2009. USENIX Association.

[10] https://blog.codecentric.de/en/2012/02/spring-data-neo4j/

ABOUT AUTHOR

G. Lawrence Paul Sundararaj received his B.Tech degree in Information Technology from Dr. Sivanthi Aditanar College of Engineering, Tiruchendur in 2012. Currently, he is pursuing M.Tech degree in Information Technology in Dr. Sivanthi Aditanar College of Engineering, Tiruchendur. His research areas of interest include Network Security, Data Mining and Knowledge Engineering.

N.Subbulakshmi received the B.E degree in Computer Science and Engineering from Jayamatha Engineering College, in 2003 and she received M.Tech in IT from Manonmaniam University, Tirunelveli, in 2005.She is an Assistant Professor in the Department of IT at Dr. Sivanthi Aditanar College of Engineering. Her research interests include Image Processing and Networking