SybilDefender: A Defense Against Sybil Attacks in Social Network

Bharati Bargot
Department of Computer Engineering
(Mumbai University)
Vidyalankar Institute of Technology
(Mumbai University)
Mumbai,India

Prof.Umesh Kulkarni
Department of Computer Engineering
(Mumbai University)
Vidyalankar Institute of Technology
(Mumbai University)
Mumbai,India

***Abstract*—Social media have gained increased usage quickly as online social network (OSN) has connected to people's everyday lives as virtual gathering places that facilitates communication. OSNs such as Facebook, Whatsapp, Twitter, Google+ and LinkedIn have hundreds of millions of daily active users. Distributed systems are vulnerable to Sybil attacks in which opponent creates many fake identities called 'Sybil Identities' and deals with running of the system or violates the system with fake information. Because of fake identities, it is vulnerable to Sybil attack. The proposed system is a mechanism that influences the network structures to defend against Sybil attacks in social network. The mechanism works based on limited number of random walks on the social graph. The system will be having algorithms such as Sybil identification algorithm and Sybil community detection algorithm and also the combination of both algorithms.**

Keywords—Social network, Sybil attack, Sybil Detection

I. Introduction

Many distributed applications and everyday services assume each participating entity controls exactly one identity. When this assumption is unverifiable the services is subject to attack. In a Sybil attack, an opponent creates a large number of fake identities also known as Sybil identities and since all Sybil identities are controlled by opponent. It can maliciously introduce a considerable number of fake opinions into the system and convert it by making decisions benefiting system itself. To defend against the attack, there have been several attempts in the form of defences or mitigations to defend against the impact of attack. Such attacks can be broadly classified into two types namely **centralized defences** and **decentralized defences.**

Recently there has been increasing interest in defending against Sybil attacks in social network [1-5]. In a social network, two user identities share a link if a relationship is established between them. Each identity is represented as a node In a social graph. To prevent the adversary from creating many fake identities, all the previous Sybil defense schemes are built upon the assumption that the number of links between Sybil nodes and honest nodes, also known as attack edges are limited. But as a result then also an opponent creates many Sybil nodes and link them in an arbitrary way. There will be small cut between honest region and Sybil region consists of all the attack edges and its removal disconnects the Sybil nodes from the rest of the graph which is influenced by previous schemes to identify Sybil nodes. Note that the solution to this problem is nontrivial because finding small cuts in a graph is NP-hard problem. To limit the number of attack edges, previous schemes assume that all the relationships in social networks are trusted and they reflect the trust relationships among those users in real world and thus an adversary cannot establish many relationships with the honest users. However, it has been shown that this assumption does not hold in some real world social networks. Some previous defense schemes can achieve good performance on small network sample but their algorithms are computationally intensive and cannot scale to networks with large node samples of online social network (OSN).The proposed system is a centralized Sybil defense mechanism, It consists of Sybil identification algorithm to identify Sybil nodes, a Sybil community detection algorithm to detect the Sybil community surrounding the Sybil nodes and two approaches to limiting the number of attack edges in online social network. The system is based in the observation that a Sybil must go through a small cut in the social network to reach the destination honest region. An honest node on contrary is not restricted and combination of two algorithms will reduces a large proportion of computation overhead.

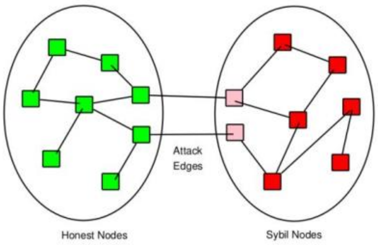

Fig. 1. Sybil attack: Nodes and attack edges

II. Literature Survey

Sybil attacks are becoming progressively severe in online social networks. An infinite number of approaches to defend against Sybil attacks have been developed in the past. Some of them are mentioned below:

1. **Sybil Limit**
   It uses mechanism of multiple random walks performed by each node. It limits the number of Sybil nodes accepted and pushes the approach to the limit. SybilLimit distinguishes Sybil nodes from honest node based on graph mixing time [2].
2. **Sybil Infer**
   It uses mechanism of Bayesian inferences on the results of random walks. Sybil Infer uses a probabilistic model defined over random walks in order to infer the extent to which a set of nodes, X, which generated such traces, is honest. It improved applicability and performance. It is less scalable and computation overhead [4].

3. **Sybil Guard**
   It uses mechanism of random walks performed by each node. It separates the social network into two regions namely, honest region and Sybil region. Sybil Guard identify Sybil node but suffers from false negatives [3].
4. **Optimal Sybil-Resilient Node Admission Control**
   It is decentralized Sybil defense scheme which relies on assumption that the social networks are random expander. It cannot effectively identify Sybil nodes on the real-world asymmetric social topologies [5].

III. Proposed System

We propose a Sybil defender, a centralized Sybil defence mechanism. It consists of different modules and different methods to defend from Sybil user. In this a user need to register for the use of social network site and authentication of user is done by user authentication module if multiple identities are not found. Our proposed system based upon the activities performed by the users in a social networking environment. User details and Sybil user's details stored in database along with their post details. The application takes into concern the user registration time, login time, activities perform and the rate at which a particular user adds friends in the system along with their activities. If all the factors are accountable then that user is considered as 'Sybil'.

System denotes the social network as a graph G containing of vertices V and edges E. There are n honest users in the social network, each with one identity, denoted as an honest node in V. There are also one or more malicious users in the social network, each with a number Sybil identities. Each Sybil identity is denoted as a Sybil node in V. A relationship between two identities in the social network is represented as an edge connecting the two corresponding nodes in G. The edges in G are undirected. The edge between an honest node and the Sybil node is represented as an attack edge. The Sybil region consists of all Sybil nodes while the honest region consists of all the honest nodes. All the Sybil nodes are controlled by an opponent. Thus, an opponent can create arbitrary edges within Sybil region.

This approach is built upon following assumptions [6]:

1. The honest region is fast mixing-Generally speaking, random walks in a fast mixing graph converge quickly to the stationary distribution.
2. One known honest node- This node is the starting point of our Sybil identification algorithm.
3. The administrator knows the social network topology.
4. The size of the Sybil region is not comparable to the size of the honest region.
5. The number of attack edges are limited.

The system will be having following algorithms:

1. Sybil Identification Algorithm to detect Sybil node.
2. Sybil Community Detection Algorithm to detect community of Sybil nodes.

Two approaches for limiting the number of attack edges:

1. Relationship rating
2. Activity Network

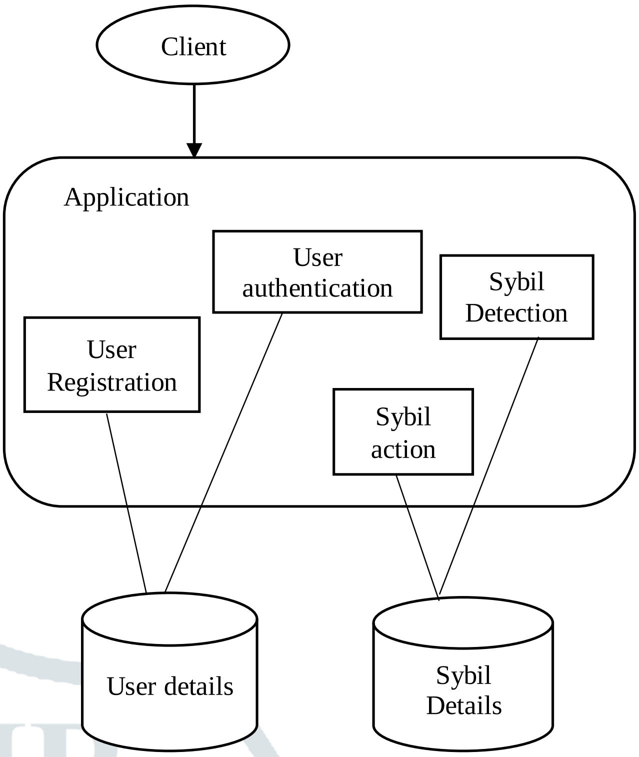

Fig. 2. System Architecture

IV. Implementation

The system consists of following algorithms:

1. Sybil Identification Algorithm
2. Sybil Community Detection Algorithm

1. Sybil Identification Algorithm:

Phase 1

1. It will take graph and one honest node as input.
2. The algorithm first performs f short random walks with length ls = log n starting from honest node h.
3. After this the known honest node and f ending node is treated as a judge node from which the algorithm sets up the criteria to identify Sybil node.
4. Now algorithm performs R random walks originating from every judge node and counts the number of nodes whose frequency is no smaller than threshold t which is a small constant.
5. The algorithm collects f +1 such value for each length l.
6. Then it compares mean and standard deviation of f+1 values and outputs a tuple <1,mean,stdDeviation >

Phase 2

1. In phase 2 the algorithm first performs random walks with initial length l originating from suspect node u.
2. The algorithm then compares the number of nodes whose frequency is not smaller than t with the mean value in tuple from algorithm 1.
3. If the former is smaller than the latter by an amount larger than stdDeviation * α consider u is Sybil and end the algorithm.

4. Otherwise the algorithm doubles l and repeats the process, until l is larger than lmax, we consider it honest and end the algorithm.

2. Sybil Community Detection Algorithm

Phase 1

1. The task of phase 1 is to estimate the needed length of the partial random walks used in phase 2.
2. Starting from an initial length l0, the algorithm performs R partial random walks originating from s and cunts the ratio of dead walks that cannot proceed before they reach the required length.
3. If this ratio is smaller than β, a threshold close to l, the algorithm doubles the current length and performs the partial random walks again. The process is repeated until the dead walk ratio is not smaller than β.
4. Then the algorithm outputs the current random walk length l.

Phase 2

1. In phase 2 it takes G, s and the estimated length l as input and outputs the Sybil community surrounding s. The reason why we need phase 2 is that not all the nodes traversed by the partial random walks in phase 1 are Sybil nodes as some walks pass the small cut and enter the honest region and we need an algorithm to select the Sybil nodes from the set of traversed nodes. To achieve this, phase 2 introduces a metric called conductance.
2. Conductance is 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'. Then the conductance of S is a/d. The conductance of set S measures the quality of cut between S and S'. The smaller the conductance is, the smaller the cut is.

   i. Phase 2 runs by first performing R partial random walks originating from the known Sybil node s, with the length decided by phase 1.

   ii. Then the algorithm sorts all the traversed nodes by their frequency in decreasing order.

   iii. Starting from the first node, which is always s, the algorithm iterates the sorted list and adds the encountered node to set S.

   iv. After all the nodes in the sorted list are examined, the algorithm records the current conductance value, starts a new iteration from the top of the list and examines each node that is not in S. This process is repeated until the conductance value stays the same at the end of two consecutive iterations.

   v. Then the algorithm outputs S as the detected Sybil community.

Two approaches for limiting the number of attack edges:

1. Relationship rating- This is one approach for limiting the number of attack edges in the network is to allow the users to rate their relationships. The users will rate their relationship by giving name to individual relationship. The relationship with name 'Sybil' will be removed from the social graph [6].
2. Activity Network- In activity network two nodes share an edge in an activity network if and only if they have interacted directly through the communication mechanisms [7] [8].

V. Conclusion

In this paper we represent the Sybil defender system which makes the actual usage of social network. The proposed system would be efficient and scalable to large social networks helps to detect Sybil identities and thwarts them for get into social site. During registration system checks for multiple individualities of new user. If identify then stop them from registration. A Sybil identification algorithm would be effectively detecting the Sybil node and Sybil community surrounding a Sybil node. System also proposing a combination of both algorithms. It uses two approaches for limiting the number of attack edges such as relationship rating and activity network.

References

[1] 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. [J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] 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.

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

[4] G. Danezis and P. Mit, "Sybilinfer: Detecting Sybil Nodes Using Social Networks," Proc. Network and Distributed System Security Symp. (NDSS), 2009.

[5] N. Tran, J. Li, L. Subramanian, and S.S. Chow, "Optimal Sybil-Resilient Node admission Control," Proc. IEEE INFOCOM, 2011. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[6] SybilDefender: A Defense Mechanism for Sybil Attacks in Large Social Networks, Wei Wei, Fengyuan Xu, Chiu C. Tan, Member, IEEE, and Qun Li, Senior Member, IEEE, VOL. 24, NO. 12, DECEMBER 2013.

[7] B. Viswanath, A. Mislove, M. Cha, and K.P. Gummadi, "On the Evolution of User Interaction in Facebook," Proc. Second ACM Workshop Online Social Networks (WOSN), 2009.

[8] C. Wilson, B. Boe, A. Sala, K.P.N. Puttaswamy, and B.Y. Zhao, "User Interactions in Social Networks and Their Implications," Proc. Fourth ACM European Conf. Computer Systems (EuroSys), 2009.

[9] E. Novak and Q. Li, "A Survey of Security and Privacy Research in Online Social networks," Technical Report WM-CS-2012-2, College of William and Mary, 2012.

[10] J.R. Douceur, "The Sybil Attack," Proc. Revised Papers First Int'lWorkshop Peer-to-Peer Systems (IPTPS '01), 2002.

[11] M. Castro, P. Druschel, A. J. Ganesh, A. I. T. Rowstron, D. S. Wallach, ―Secure routing for structured peer-to-peer overlay networks, ‖ in Proc. of USENIX OSDI, 2002. Article (CrossRef Link)

[12] F. Lesueur, L. Me, V. V. T. Tong, ―A Sybilproof distributed identity management for P2P networks, ‖ in Proc. of IEEE ISCC, pp. 246-253, 2008. Article (CrossRef Link)

[13] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove, "An analysis of social network-based Sybil defenses," in SIGCOMM, 2010.

[14] Twitter vulnerability allows cyber criminals to spread spam. http://www.one.com/en/web-hostingnews/website/twitter-vulnerability-allows-/cybercriminals-to-spread-spam-links\$800076628.htm.

[15] Twitter accounts spreading malicious code. http://www.net-security.org/malware_news. php?id=1554

[16] J.Newsome,E.Shi,D.Song,andA.Perrig. TheSybilattack in sensor networks: Analysis & defenses. In ACM/IEEE IPSN, 2004.