Volume 6, Issue 10, October 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Defense Against Sybil Attacks in Social Networks

Kamble Madhuri D.
P.G. Student, Dept. of Computer Engineering,
Terna Engineering College, Nerul, Navi Mumbai,
Maharashtra, India

Prof. Gaikwad Ujwala V.
Asst. Prof., Dept. of Computer Engineering,
Terna Engineering College, Nerul, Navi Mumbai,
Maharashtra, India

Abstract- A social network is a system which is a structure made up of actors such as individuals or organizations, and ties between these actors such as interactions, relationships, and connections. Formally, in the literature, this is almost always represented as a graph which we refer to as the social graph. The nodes of such a graph represents an actor and the edges represent ties between those actors. An online social network has a representation of a user (usually a profile) and his or her social links, although other services are often incorporated. These systems are vulnerable to Sybil attacks, which affects the performance of the system. The Sybil identities can "suppress" the honest identities in a variety of tasks, including online content ranking, DHT routing, file sharing, reputation systems, and Byzantine failure defences. So to protect network from such attacks so much research has done. The proposed system is a mechanism that is leverages the network topologies 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. We propose two approaches to limiting the number of attack edges in online social networks.

Keywords- Social network, Sybil, Attack, Community, Detection

I. INTRODUCTION

Distributed systems are vulnerable to Sybil attacks [1][2], in which an adversary creates many bogus identities, called Sybil identities, and compromises the running of the system or pollutes the system with fake information. Sybil attacks can also be defending by assuming the existence of trusted authorities which can limit the introduction of fake identities by requiring the users to provide some high level credentials like social security number or by requiring payment or by providing more security methods but such requirements will prevent users from accepting these systems, as they impose additional burden on users.

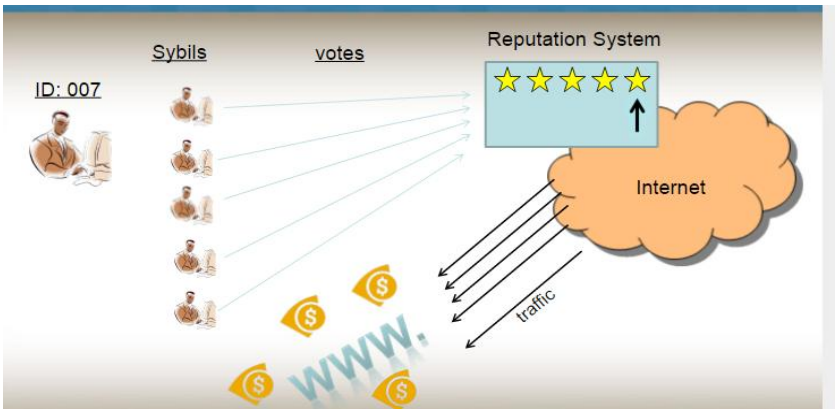

Figure1: schema of Sybil attack

Figure 1 shows schema of sybil attack. Recently there has been increasing interest in defending against Sybil attacks [3][4][5][6][7] in social networks. 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 the social graph. To prevent the adversary from creating many Sybil identities, all the previous Sybil defense schemes are built upon the assumption that the number of links between the Sybil nodes and the honest nodes, also known as attack edges, is limited. But as a result then also an adversary creates many Sybil nodes and link them in an arbitrary way, there will be a small cut between the honest region and the Sybil region. The small cut consists of all the attack edges and its removal disconnects the Sybil nodes from the rest of the graph, which is leveraged by previous schemes to identify the Sybil nodes. Note that the solution to this problem is nontrivial, because finding small cuts in a graph is an 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 the 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

In the past few years, online social networks have gained great popularity and are among the most frequently visited sites on the web. The large sizes of these networks require that any scheme aiming to defend against Sybil attacks in online social networks should be efficient and scalable. Some previous 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.

The proposed system is a centralized Sybil defense mechanism. It consists of a Sybil identification algorithm to identify Sybil nodes, a Sybil community detection algorithm to detect the Sybil community surrounding a Sybil node. And two approaches to limiting the number of attack edges in online social network. The system is based on observation that a Sybil node must go through a small cut in the social graph to reach the honest region. an honest node on contrary is not restricted.

And also the Combo algorithm that combines the Sybil identification algorithm and Sybil community detection which will reduces a large portion of computation overhead.

II. LITERATURE SURVEY

Sybil attacks are becoming increasingly serious in online social networks. A number of schemes to defend against Sybil attacks have been developed in the past.

A) SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks, 2008[4]

- Here it assumes that Non-Sybil region is fast mixing.
- Algorithm used for multiple random walks performed by each node.
- Varying number of random walks and walk length.
- Disadvantages -
  - Can identify one sybil node at a time.
  - Not scalable for large number of node.

B) SybilGuard: Defending against Sybil Attacks via Social Networks, 2006[5]

SybilGuard design leverages the following three important facts to bound the number of Sybil nodes:

- Social networks tend to be fast mixing, which necessarily means that subsets of honest nodes have good connectivity to the rest of the social network.
- Too many Sybil nodes (compared to the number of attack edges) disrupts the fast mixing property.
- The verifier is itself an honest node, which breaks symmetry.
- Disadvantages-
  - Can identify one sybil node at a time.
  - Not scalable for large number of node.

C) SybilInfer: Detecting Sybil Nodes Using Social Network, 2009[6]

- It assumes that Non-Sybil region is fast mixing, modified walks are fast mixing.
- Bayesian inference approach that assigns a Sybil probability, indicating the degree of certainty, to each node in network.
- Disadvantages-
  - High computation overhead.
  - Less scalable.

D) Optimal Sybil- Resilient Node Admission Control, 2011[7]

- Decentralized Sybil defense scheme.
- Relies on assumption that the social networks are random expander.
- Disadvantages-
  - High false positive and negative rates.
  - Cannot effectively identify Sybil nodes on the real-world asymmetric social topologies.

Analysis of previous work

Table1: Comparison

System	Mechanism Used	Algorithm Principle	Merits	Demerits
SybilGuard	Random walk performed by each node	SybilGuard separates the social network into two regions, namely, honest and Sybil.	Identify Sybil nodes.	Suffers from high false Negatives.
SybilLimit	Multiple random walks performed by each node	In SybilLimit, a tail is defined as the last edge of a random route. Verifying Suspect in SybilLimit is different	Improved version of SybilGuard, SybilLimit	Fail to defend Sybil nodes

		from that in SybilGuard.	limits the attack.	
SybilInfer	Bayesian inference on the results of the random walks	SybilInfer first transforms social network *G* into new graph *G'* for the stationary distribution of *G'* to be uniform.	Improved applicability and performance	Less scalable and computation overhead
GateKeeper	Bootstrap phase and node admission based on tickets	The Gatekeeper develops sub-graph *H* within *G* in a breadth-first (BF) manner. *H* begins from a root, selected through random walks.	significantly limits the number of Sybil's admitted per attack edge to a small value even in the face of a large number of attack edges	Cannot effectively identify Sybil nodes on the real-world asymmetric social topologies.

III. PROPOSED SYSTEM

System denotes the social network as a graph G consisting 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 of 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. System names the edge between a Sybil node and an honest node an attack edge. The Sybil region consists of all the Sybil nodes, while the honest region consists of all the honest nodes. All the Sybil nodes are controlled by an adversary. Thus, the adversary can create arbitrary edges within the Sybil region.

- This approach is built upon following assumptions [8]:
  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 is 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

IV. METHODOLOGY

- The system consists of following algorithms:
  1. Sybil Identification algorithm
  2. Sybil Community algorithm

1. Sybil Identification algorithm

- Phase1
  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 originating 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 computes mean and standard deviation of f+1 values and outputs a tuple<l, mean, stdDeviation>

- Phase2
  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 no smaller than t with the mean value in tuple from alg1
  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. If u is still not identified as Sybil when the value of l reaches lmax, we consider it honest and end the algorithm.

2. Sybil Community Detection algorithm

- Phase1
  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 counts the ratio of dead walks, which are the walks that cannot proceed before they reach the required length.
  3) If this ratio is smaller than β, a threshold close to 1, the algorithm doubles the current length and performs the partial random walks again. This process is repeated until the dead walk ratio is no smaller than β.
  4) Then, the algorithm outputs the current random walk length l.

- Phase2
  - 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 leverages a metric called conductance.
  - 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 a set S measures the quality of the cut between S and S': the smaller the conductance is, the smaller the cut is.
  1) Phase 2 runs by first performing R partial random walks originating from the known Sybil node s, with the length decided by phase 1.
  2) Then, the algorithm sorts all the traversed nodes by their frequency in decreasing order.
  3) Starting from the first node, which is always s, the algorithm iterates the sorted list and adds the encountered node to set S if doing so does not increase the conductance of S.
  4) 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.
  5) 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 the stars or name to individual relationship. The relationship with lowest stars will be removed from the social graph.

2) Activity network

In activity network [9] [10] two nodes share an edge in an activity network if and only if they have interacted directly through the communication mechanisms or applications provided by the social network. In other words, a social network is transformed into an activity network by removing the weak connections with no user activity. If the Sybil defense schemes leverage the topologies of the activity networks, the number of attack edges an adversary can create can be further limited.

V. CONCLUSION

The proposed system would be efficient and scalable to large social networks, a scheme that leverages network topologies to defend against Sybil attacks in large social networks. It consists of a Sybil identification algorithm, a Sybil community detection algorithm. Sybil identification algorithm would be effectively detecting the Sybil node and Sybil community detection algorithm would effectively detect Sybil community surrounding a Sybil node. System also proposing a combination of Sybil identification algorithm and Sybil community detection algorithm. And also two approaches for limiting the number of attack edges such as relationship rating and activity network.

REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

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

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

[7] N. Tran, J. Li, L. Subramanian, and S.S. Chow, "*Optimal Sybil- Resilient Node admission Control*," Proc. IEEE INFOCOM, 2011.

[8] *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.

[9] 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.

[10] 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.

