IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 6, Ver. II (Nov – Dec. 2014), PP 01-06
www.iosrjournals.org

Intent Search and Centralized Sybil Defence Mechanism for Social Network

Julia George
(*Dept. of Computer Science and Engineering, Caarmel Engineering College, Pathanamthitta, Kerala, India)*

Abstract: *Sybil attacks are the major problems occurred in the distributed systems without trusted identities. It occur when the one-to-one relationship between a node and its identity is violated. This is occurred by an attacker creates multiple pseudonymous identities and pretends to be multiple, distinct nodes in the system. Intent search is another important function in social network. It mostly depends on surrounding text features which leads to noisy search result.*

In this paper a new Sybil defence mechanism is introduced that is Sybil protector, in order to defend against attacks in social networks. It is efficient and scalable to large social network. Network topologies are used by the Sybil protector to defend against Sybil attacks in social networks. It can effectively identify the Sybil nodes and detect the Sybil community around a Sybil node. And also limits the number of attack edges in online social networks. Here also included a novel internet image search which only requires one-click user feedback. For that expanded keywords are used to create positive example images and to include more relevant images it enlarge the image pool. It consist of four steps.1) Adaptive similarity.2) Keyword expansion. 3) Visual query expansion. Intent image search is very effective because it has extremely simple user interface.

IndexTerms: *Sybil attack, Sybil identity, social networks, attack edge, keyword expansion.*

I. Introduction

Decentralized distributed systems are vulnerable to malicious attacks where an adversary pretends to have multiple identities. These type of attack is known as Sybil attack is known as Sybil attack and such identities are known as Sybil identities. Sybil identities suppress by key functions of the honest identities in the web-based distributed systems. Sybil attack can highly influence the working of open membership systems such as eBay, Facebook, BitTorrent etc.

Traditional approach for preventing there Sybil attacks relay on trusted authorities, when certify identities. But asking users to present credentials such as social security numbers or other trusted identities will affects the success of open membership systems. However, these requirements adds extra burdens on users, it badly affects users intention to join these systems.

Recently, there has been lots of research to mitigate Sybil attacks in social networks. Monitoring historical behavior, of an identity is insufficient to prevent Sybil attacks because they pretends to be nice and can launch attacks at any time. There are two types of Sybil defense mechanisms, centralized and decentralized. Defending against Sybil attacks using a centralized approach is much harder. One simplest approach is to bind nodes to IP addresses. And another computation puzzles which needs human efforts, such as CAPTCHAS. But both of them can only provide limited protection against attacks. Note that an attacks can easily harvest IP address, while CAPTCHAs can be solved by reposting it on attacker's website where other nodes trying access to that particular web site.

Two users in a social network establish a relationship between them, they share a link. Each honest users is known as a node or honest identity. Attack edges are the links that connects the Sybil nodes with the honest nodes. Previous Sybil defense mechanisms are based on the assumption that the number of attack edges are limited. Small cut is the region between honest and Sybil region, and it consist of attack edges. By removing these attack edges Sybil nodes are disconnected from the graph.In this paper, a centralized Sybil defense mechanism is proposed. It consist of two parts Sybil identification algorithm and a Sybil community detection algorithm.

II. Related Works

Utilizing social network topologies is a way to defend Sybil attacks. Sybil Guard [1] is a decentralized algorithm which limits the Sybil attacks. And also limits attacks from botnets that is from outside the system. This protocol states that the number of attacking edge does not depend on number of Sybil nodes. In a social network graph, the nodes or vertices represents the identities in the distributed system and the edges (undirected) represents to human trusted relations from the real world. Sybil guard finds a small quotient cut in the graph i.e. a small set of attack edges, its removal will disconnects a large number of Sybil identities. In other words fast mixing social networks do not have this type of cutes.to limit the size of attack Sybil guard uses the small quotient cut. In Sybil guard honest node v is called the verifier and node s is called the suspect. Node v decides whether to accept node s or not.it suffer due to high false negatives. And in each attack $o(\sqrt{n}/\log n)$ Sybil identities remain undetected. Since Sybil guard is better than previous approaches suffer from two limitations. First is it allows $o(\sqrt{n} \log n)$Sybil node on each attack edge i.e. for a million-node social network there are 2000 Sybil nodes are accepted per attack edge. Second is an assumption on which Sybil guard relies on social networks are fast mixing but in real world that had never been violated.

Sybil limit [2] protocol limits the number of Sybil identities accepted and it is also near optimal. It allows $o(\log n)$ Sybil nodes per attack edge. It accepts only 10 Sybil nodes per attack edge. It is 200 times of Sybil Guard. In other words adversary needs 100,000 trust relations with honest identities compared to Sybil Guard, it needs only 500 trust relations.

Multiple walks accepting fewer Sybil nodes per attacks edge in Sybil limit is improved than Sybil Guard's bound without compromising on other features to Sybil Guard, Sybil limit archives these improvements. In order to drive improvement in Sybil limit multiple novel techniques are combined together. For performing many short random routes 1) it leverages multiple non-independent instances of the random route protocol ii) Instead of identities intersections on edges are exploited, iii) to deal with escaping tails of the verifier novel balanced conditions are used and to safety estimate r, novel bench marking technique is used . Results from Sybil limit conformed the fast mixing property of real world social network and the main assumption behind this approach is validated.

Centralized Protocol is used in Sybil Infer [3] and it assumes knowledge regarding social graph. It assign a probability of being Sybil identity to each node by using a Bayesian inference technique. Unlike Sybil Limit and Sybil Guard, o the number of Sybil nodes accepted provide any analytical bound by Sybil Infer. Because of the use of machine learning techniques and their approach to the security issues, Sybil Infer is very significant.

Decentralized Sybil detection protocol is used in GateKeeper [4] .It is improved than Sybil limit. A variant of ticket distribution algorithm in sumUp [5] is used by GateKeeper for multiple random nodes in the graph to identify Sybil. It allows $o(\log k)$ Sybil nodes per attack edge. On a random expander graph this protocol gives a factor of $o(\log h)$. To perform identity admission control in a decentralized way. It uses an improved version of the ticket distribution algorithm SumUp. From different taken vantage points and it executes the ticket distribution algorithm and by combining these result which perform decentralized admission control.

III. System Model

Sybil detection schemes are designed on identity based social network. Each exploiter have a single identity or node. They establish relationship links that they recognize in the system, thereby creating a social network. Social network is represented as a graph G which consist of vertex V and edges E. In a social network there are n honest users, with one identity denoted as an honest identity as v. Honest identity follows the protocol. Between all nodes in the system there exist an undirected graph. An undirected edge between two honest nodes reflects the strong social connections between those nodes in the real world i.e. if two users are friends they are connected by an edge. Among all nodes the knowledge of the social graph is established. On a social graph each honest identity knows its immediate neighbors and may not know the rest of the system

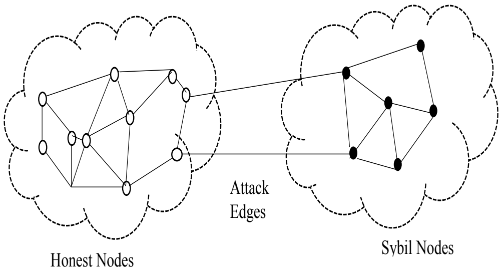

Fig 1. Illustration of Honest nodes, Sybil nodes and attack edges between them.

The system also consist of one or more malicious users, each with a number of distinct identities. These identities are known as Sybil identities i.e. all the identities created by a Sybil are referred as Sybil identities. Nodes are also called identity and in this paper the terms "identity" and "node" are used interchangeably. Sybil nodes behave arbitrarily. All Sybil identities may collude, and are controlled by an adversary.

Sybil nodes can create link with honest node by successfully fools an honest node. This type of link or edge that exist between the Sybil identities is called an attack edge. Honest nodes are combined to form honest region while Sybil nodes re combined to form Sybil region All Sybil identities and honest identities in the system together to form a social network. The Sybil attack has full knowledge of the entire social network .The goal of the Sybil Protector is to detect Sybil nodes and honest node with high accuracy. And also detect the Sybil community.

Following assumptions are used in Sybil Protector

The honest region is fast mixing : The honest region of the social network are fast mixing or density connected, which means random walks in the honest region quickly reach a state of stationary distribution.

At least one trusted node: These should be at one known identity in the social network .In Sybilidentification algorithm this node is taken as the starting node.

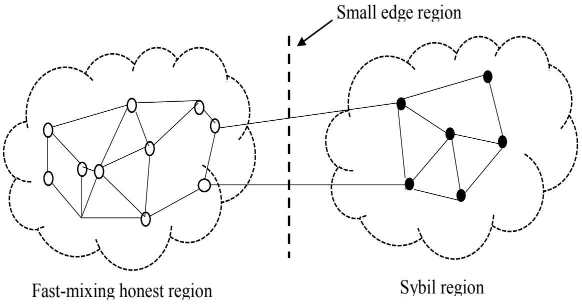

Fig 2. Sybil detection relies on the small edge cut between the fast mixing honest region and the Sybil region

The administrator has full knowledge about the network topology: That means Sybil Protector is a centralized Sybil defense mechanism considering social network is controlled by an administrator. So administrator can take charge of reducing Sybil attacks.

Size of the honest region and Sybil region are not comparable: Assume that adversary cannot establish such many Sybil nodes as existing social networks. Because of during signing up a new honest always includes some personal information verify an email address and solving CAPTACHAs.

Limited number of attack edge: There will be a disproportionality small cut between Sybil region and honest region when the adversary creates many Sybil nodes. Fast mixing property was distributed because of the existence of a small cut: the mixing between honest identities and Sybil identities are slow while mixing between honest identities are fast.

IV. Sybil Protector Design

Sybil Protector includes three steps: an honest node detection algorithm, Sybil identification algorithm and Sybil community identification algorithm. The steps can be used together to mitigate Sybil attacks. Section IV –A explains the working of highest node identification algorithm which determine the degree of each node. Section IV-B explains the working of Sybil identification algorithm, it determines whether an identity is Sybil or not. Then is section IV-C in that Sybil community identification algorithm is explained, which shows how to effectively recognize the Sybil community around a Sybil identity. The community detection is impractical by simply analyzing all the identities in the social graph that's why third algorithm is used. To identify Sybil's, all algorithms are based on the assumption that there is only limited number of attack edges.

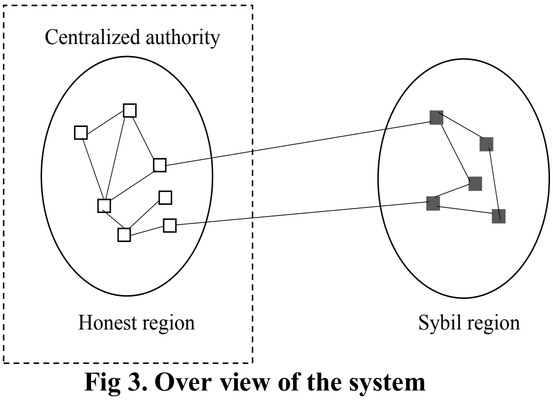

Fig 3. Over view of the system

A. Highest Honest Node Identification Algorithm

In this subsection honest nodes are identified by using social graph *G (V, E)* where inputs honest (user) node *h*, and a friend node *f* as inputs and outputs whether f is honest. In this algorithm there are two sets, friends and users. *l user* is the login user*, f count* is the friends counts and *n* is the total number of users. Initially the count of *f* is set to zero.

Algorithm 1 GetHigestHonestNode(G,l user)

```
l user = login user
f count = 0
honest graph =0
friends { }, user{ }
n = total number of users
fori = 1 to ndo
if check friends in user list then
  friends(k) = user(i)
  k++
  honest graph. add user(i)
end if
end for
get honest nodes
l vertex =login user
f vertex = friend
count = 0
for i=1 to k do
friends { friends [ i ]}
friends of friends [c] =friends  [ i]
fn.get honest nodes( friends offriends{ })
f vertex = vertex( friends [ ])
honest graph. Add (l vertex, f vertex, i)
  n friend list (i) =friend (i)
  if count<= 5000
  count ++
  get honest nodes (new friends { })
end if
end sub
end for
```

Given a social graph G(V,E) and a known honest identity i.e. login user denoted as *l user*. Initially set friends count, *f count* as zero. And *honest graph* as zero. Then check the login user's friends list and generate a graph for the login user. Next is to find the honest nodes. For that friends of friends list is taken. Call a function *get honest node* and in that function *friends of friends list* is called recursively And check their friends up to 5000.If all the nodes are added in the *f vertex* they are considered as the honest nodes. The nodes which are not added in the list have a possibility to become a Sybil node.

B. Sybil Identification Algorithm

The suspicion of Sybil identification algorithm is that, in between honest region and Sybil region there is small cut. The random walk starting from a Sybil identity tend to get "pinned" into the Sybil region. The size of the Sybil region and the honest region are not comparable. The number of identities traversed by the random walks starting from a Sybil identity is smaller than the number of identities traversed by the random walks starting from honest identity. Frequency of a node is defined by the number of times a particular node being traversed by a set of random walks. And it is denoted as *frequency*. Note that multiple times one node can be traversed in the same direction.

Algorithm 2 SybilIdentification(G,highest honest nodes)

```
frequency, wl,Sybil { },Sybil count=0
for i = 1 to k do
forj = 1 to count do
if ( friend (i). identity== n.friend(j) .identity) or( non-repliedmsg> 100) then
frequency++;
```

```
Sybil {i}
end if
wl ++
end for
if frequency >mean of friends countthen
  Sybil count++
 else
wl++
end if
end for
output Sybil count
```

In Sybil identification algorithm social graph G(V,E) is taken and also get the output of the previous algorithm. Firstly initialize frequency, walk length which is denoted as *wl*, Sybil list and *Sybil count*as zero. First take nearest friends list and then take their friends list. Then check their similarity of their given identities. If identities are similar they can be considered as Sybil. Another criteria checked is the non-replied messages if it is greater than 100 they can be also considered as Sybil's. Then check the frequency i.e. how many times a particular node visits an honest nodes profile according to the number of times frequency count also increases check the walk length of each node. From all these list compared together to get the Sybil count.

C. Sybil Community Identification

Algorithm 3 SybilCommunityIdentification( )

```
traversed list { }, Sybil community { }
fori =1 to Sybil count do
  for j =1 to count do
     traverse list. Add friends (j)
  end for
end for
s.count= 0 do
s.count = existence (sybil)
for i =traverse list.first() to traverse list last () do
if node i∉sybil then
continue
else
Sybil community = Sybil (i)
end if
end for
while s.count<Sybil count
output Sybil community
```

In Sybil community identification, take the nodes from previous count and traverse the graph. Then set *s.count*as zero. Then check the existence of the Sybil. Traverse the graph from first node to the last node. If the node is Sybil, they are added to the Sybil community.

V. Evaluation

In this experiment a social network model is created. In that Sybil nodes and honest nodes are also created. Sybil communities are also build by making relationship among Sybil nodes. The topology and social graph of this network is used for the evaluation. It is a random network with no particular format and Sybil regions also have different arbitrary structures. First an honest region and then a Sybil region is created and connect this Sybil region to the social network. Initially adversary launch attack to a small number of nods then fastly increase the attack edges. So first select some nodes from the data set randomly then increase attack edge.

From this created social network experiments are carried out by varying the number of honest nodes and the number of Sybil nodes. By summarizing this experiment it is found out that 75% of Sybil identification are done effectively comparing with the previous scheme Sybil Protector detects more Sybil nodes and their surrounding community. Sybil protector identify the small cut exist between the Sybil region and honest region effectively. And it can effectively identify Sybil identities when the number of Sybil nodes commenced by each attack edge comes up the theoretical lower bound.

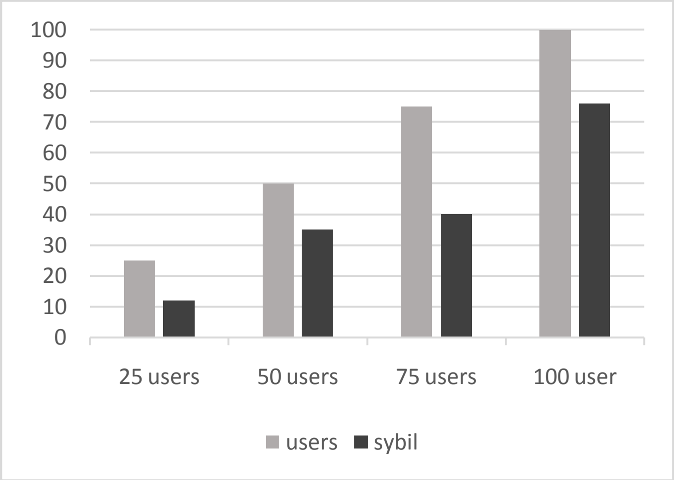

Fig 5. Comparison of Sybil detection on varying number of users in asocial network

Here I try to introduce a better scheme for intent search images in social networks. This scheme needs only one-click user feedback. In existing system there is only i) Adaptive similarity ii) Keyword expansion iii) Visual query expansion. More than that in this proposed system color scheme identification, synonyms, height and weight image clustering is used. First in adaptive similarity, predefined adaptive weight categories are used to categorize query image expansion. In keyword expansion visual content of the query image are referred which is not used in traditional text based expansion. In visual query expansion images which are similar to query images are found.

Designed and followed a set of features that are good in describing visual contend of images and which are effective in their storage and computational complexity. Existing features used are i) SHIFT [13] ii) Histogram of Gradient [12]. According to the query images and the keyword provided by the user 'the demand of the user is further caught in two aspects: i) checking query keywords representing users intention more effectively ii) And at the same time finding a group of images which are both semantically and visually consistent with the input image.

Attention guided [11] color signature is used in this system.Color signature describes the color composition of an image. Clustering colors of pixels in the cluster centers and LAB color space are done and then their relative proportions are taken and they are considered as the signature. In order to compute saliency map of the image an attention detector is used.

VI. Conclusion

This paper presents Sybil Protector a centralized protocol for identifying sibyls and Sybil community. Sybil protector trust on the underlying properties of the social networks such as i) fast mixing property of the honest region ii) attacker create many nodes but only few attack edges. This system can effectively identify the Sybil identities and it can correctly recognize the Sybil community around a Sybil node with different structures and size.As future work, accuracy of the Sybil detection scheme can be improved by modifying the algorithm. And can also increase the accuracy by tracing the IP address of each nodes. Using intent search it is possible for scale image search by both visual and text content, which only needs one-click user feedback.

References

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

[2] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. Sybillimit: A near optimal social network defense against sybil attacks. In IEEE Symposium

[3] G. Danezis and P. Mittal, "SybilInfer: Detecting Sybil Nodes using Social Networks," in NDSS, 2009.K.Elissa, "Titleofpaperifknown," unpublished.

[4] N. Tran, J. Li, L. Subramanian, and S. S. Chow, "Optimal sybil-resilient node admission control," in INFOCOM, 2011.

[5] N. Tran, B. Min, J. Li, and L. Subramanian, "Sybil-Resilient Online Content Voting," in NSDI, 2009.

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

[7] SybilDefender: Defend Against Sybil Attacks in Large Social Networks J. R. Douceur. The sybil attack. In IPTPS, 2002.

[8] IntentSearch:Capturing User Intention for One-Click Internet Image SearchY. Ke, X. Tang, and F. Jing, "The design of high-level featuresfor photo quality assessment," in Proc. IEEE Int'l Conf. ComputerVision and Pattern Recognition, 2006.Y

[9] Rubner, L. Guibas, and C. Tomasi, "The earth movers distance, multi-dimensional scaling, and color-based image retrieval," in Proc. the ARPA Image Understanding Workshop, 1997.N.

[10] Dalal and B. Triggs, "Histograms of oriented gradients for human detection," inProc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2005.

[11] D. Lowe, "Distinctive image features from scale-invariant keypoints,"International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004..