International Journal of Science and Research (IJSR)
ISSN (Online):
Index Copernicus Value (2016): 6.14 | Impact Factor (2015): 6.391

# Vote Credence: Social Network Sybil Defence by User Behaviour

**Anagha P B[1], Janitha Krishnan[2]**

[1]AWH Engineering College, Calicut University, Department of Computer science & Engineering, Kuttikkatoor, Kozhikode, India

[2]AWH Engineering College, Calicut University, Department of Computer science & Engineering, Kuttikkatoor, Kozhikode, India

**Abstract:** ***Online Social Network (OSN) typically consists of different type of users and it allows efficient communication between persons. Due to open nature of OSN, it's always vulnerable to different type of attacks. Attackers can create fake identities in OSN to degrade the reputation of the system. Such fake identities are called Sybil's. Vote Credence is a Sybil detecting and blocking mechanism. Vote Credence uses both the concept of friend invitation graph and user behaviour analyser logic to defend against Sybil. It is a voting based Sybil defending system. It efficiently identifies the Sybil's that propagate malicious messages by a message filtering system included in the user behaviour analyser. This system helps us to block the entire profile of Sybil.***

**Keywords**: Online Social Network, Sybil, Sybil Attack, Sybil Detection, Message filtering

## 1. Introduction

Recently Online Social Network (OSN) has become very popular to make connections within people in real life. There are different online network services such as Facebook, Twitter, and LinkedIn. These social networks are unprotected because of many reasons. Because of its open nature, different type of attackers tries to infiltrate into the OSN system. A particular type of attack against the OSN is considered in this paper. It is called Sybil attack. Sybil's are the multiple fake identities or profiles in Online Social Network.

Sybil may affect the OSN and profiles of persons within the OSN very seriously. They make changes in the reputation and popularity of persons those who are active in the online social network. Particularly Sybil's are becoming a threat to all the networking sites. This paper deals with a mechanism to defend against the Sybil attack by considering user level activities.

Many of the prior Sybil defending mechanism is developed by considering only the friend request/confirm behaviour of the user. They don't consider what type of messages they sent to other people and also it doesn't consider whether the message consisting of any vulgar content. This can be considered as the limitation of the prior work [8].

In this paper we also explore the negative distrust relationships which are in the form of rejected friend requests. This is based on the assumption that Sybil's have many negative relationships (Many rejected friend requests) than positive ones. But we can't apply this feature because Sybil can create many negative trust relationships between Sybil's. Sybil can use this type of relationship to provide an unclear picture about the Sybil to the Sybil detector.

To identify the fake relationship we can make a data structure called friend invitation graph. We can model this graph by considering the persons in the graph as nodes, and the relationship between the persons as edges of the graph. In Vote Credence, we say that a node B casts a (positive/negative) vote on node A if B accepts or rejects the request from A. After that Vote credence identifies some suspicious profiles which are rejected by many persons. Then the system to defend against Sybil applies the text filtering and media content filtering to filter out the messages that contain substandard contents. Based on the analysis of text filtering and media content filtering vote obtained by each of the users in the OSN will be calculated. Vote Credence aggregates the vote obtained by each users in the OSN. That is we can evaluate the probability of being a real user. During this analysis Vote Credence penalize the vote from suspected nodes by blocking the entire profile of the persons.

The rest of the paper is organized as follows. Section 2 presents the literature survey, followed by Sybil detection in vote credence. Text filtering is presented in section 3.2.The media content filtering that can be introduced on this system can be included in section 3.3 followed by conclusion and future work of this paper

## 2. Literature Survey

Literature survey deals with the related techniques which contribute to the development of Vote Credence system. There are a number of works related on different types of Sybil defending mechanism. Some of them[1] analysis Sybil attack and identity clone attack based on different things, such as attack impacts ,pre-requirements and network topology. There are also some Sybil defending schemes which can be based on social graph based approaches [2] [3].

Sybil Guard [2], is a novel decentralized protocol, it is based on the assumption that social network are fast mixing. The basic idea behind [2] is that an honest node (called the verifier) decides whether or not to accept another node (called the suspect).The verifier only accepts a suspect whose random route intersects with the verifier's random route. Some limitations of [2] are identified as it does not limit the Sybil nodes. Some other protocols [3] [4] are available to defend against Sybil's. Sybil Limit [3] is also a protocol

presented by Haifeng Yu and Michael Kaminski, that leverage the same insight as [2] but offers dramatically improved and near-optimal guarantees. Sybil Limit adopts a similar system model and attack model as presented in [2]. Some limitations of [3] are identified as it require undirected un-weighted graph and it is favourable to newcomers with many links and unfavourable to those with few links.

Sybil Infer [4] is an algorithm presented by George Danezis and Prateek Mittal for labelling nodes in a social network as honest users or Sybil's controlled by an adversary. Similar to Sybil Infer [4], both [2] and [3] protocols exploit the fact that a Sybil attack disrupts the fast mixing property of the social network topology guarantees. An important difference between Sybil Infer [4] and presentation of Haifeng Yu and Phillip B. Gibbons is that the former is not sensitive to the degree of the attacker nodes, [3] provides very weak guarantees when high degree (e.g. degree 10) nodes are compromised.

Sum Up [5] is a Sybil resilient online content rating system that leverage trust networks among users to defend against Sybil attacks with strong security guarantees. In the previous studies of decentralised protocol [2] Non-Sybil region is fast mixing. In Nguyen Tran proposal [5] Non-Sybil region is fast mixing. Viswanath [6] study on Sybil defence schemes identifies various features of these Sybil defending schemes. Gate keeper [7] is a distributed Sybil-resilient admission control protocol. This work [7] uses an improved version of the ticket distribution algorithm [5] to perform node admission control in a decentralized fashion. It uses a similar system model and threat model as those used in previous systems Sybil Limit [3], and Sybil Guard [2]. The key idea of the study [7] is to perform distributed ticket distribution from multiple ticket sources. The limitation in previous scheme [5] is concerned with a single source not being able to reach the vast majority of honest nodes [7], an honest node not reachable from one source may be reached by other sources.

Vote Trust [8] further explores the negative distrust relationships (e.g., in the form of rejected friend requests) among users.

## 3. Sybil Detection

This section describes the individual Sybil detection in the OSN. In the design of Vote Credence, it considers the Sybil detection as a vote aggregation problem. Votes are mainly collected from real users only. It is guaranteed by Vote Credence. In Vote Credence, each node has an important feature**: Global acceptance rate** *(GAR)* is the fraction of positive votes that Vote Credence aggregates for a single user, indicating the probability that user is accepted by real users. The profiles of users with low global acceptance rate (e.g., below a certain **threshold** *δf)* are identified for further analysis. Vote Credence also check the messages send by the persons to identify whether it contains any bad messages or not. Both text message and images send by the persons are considered. In this design both text and image messages that containing substandard contents will be blocked. The Sybil's will be identified on the basis of how many messages are blocked by each person.

### 3.1 System Architecture

Vote Credence consists of following filtering layers, Text filtering and Media Content Filtering.

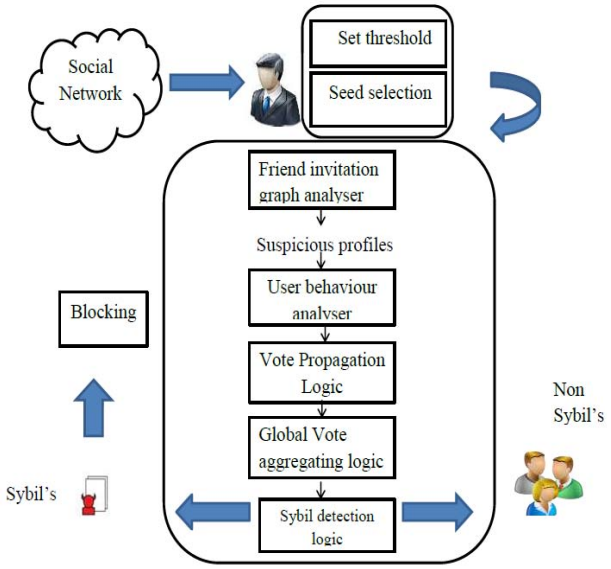


**Figure 1:** Vote Credence Architecture

Architecture starts from the online social network. Every OSN consist of an admin. Basically admin have different roles in OSN. In Vote Credence, some sort of settlement of threshold and seed selection can be performed by the admin

**Selecting Trusted Seeds:** The goal of seed selection is to find real users that will be the most useful in identifying other real users. A heuristic for selecting seeds is to give preference to those from which trust can be propagated to many other real users. This selection of seeds can be performed by both admin and users in the OSN. Admin can send mails that contain the link of signup to the vote credence system .It can be send to some trusted users. These users obtain the privilege to sign into the system and they can also have the ability to send this mail to other users. We are assuming that real users prefer to send requests to their real-life acquaintances.

**Friend invitation graph analyser:** Friend invitation graph is a data structure used by vote credence. It is modelled by the system by considering the persons in the social network as nodes and their relationship as edges. In Vote Credence, we say that a node B casts a (positive/negative) vote on node A if B accepts or rejects the request from A. After that Vote credence identifies some suspicious profiles which are rejected by many persons. Based on the acceptance of friend requests, acceptance rate (ACR) of each node (user) will be calculated. It will be calculated as number of Accepted request's to the sum of number of accepted requests and number of rejected requests

Nodes with low acceptance rate (Below the value 0.5) will be identifies as suspicious profiles, and it will be given to further analysis by user behaviour analyser module. We are

considering the 50% activity of every person, so the value 0.5 can be considered.

**User Behaviour analyser:** This module analyse the behaviour of users in the system. Mainly the messages send by the persons are considered. Both text and image are considered. This module consists of text content filtering and media content filtering to filter out the messages that containing bad contents. These filtered messages help us to determine Global acceptance rate. Global acceptance rate calculation will be done by the count of messages that is been blocked. Message post rate (MPR) of each person helps us to determine the behaviour of every person. It can be calculated as number of successful messages to the sum of number of successful messages and number of failed messages.

**Vote Propagation:** After identifying some suspicious profiles by friend invitation graph analyser user behaviour analyser can detect the Sybil's by considering the count of the trusted friends in each individual profile. The trust is identified by number of friends having acceptance rate above to the total no of friends. This propagated trust score (PTS) can be calculated as count of no of friends having acceptance rate greater than 0.5 to the total no of friends.

**Global Vote Aggregating:** Global acceptance rate of each user can be calculated as the average of these three.

GAR = (ACR+MPR+PTR)/3

### 3.2. Text Content Filtering

Text content filtering is performed to prune the messages that containing bad contents. The basic steps behind the text content filtering are given below.
1)Creating a word vector
2)Checking the category of the message
3)Calculate the filter value of message, by comparing the word vector with knowledge repository for abusive words.
4)Recalculate the trust score based on the filter value.

### 3.3 Media Content Filtering

Media content filtering is performed to prune the image messages that contain vulgar contents. Initially we are considering the entire image. Then human parts are extracted from that image. Basically vote credence analyses what amount of body part are exposed in that images. If this value is above some threshold that images will be blocked. The users in the vote credence system will be blocked by the count of messages that is been blocked. The basic steps behind the text content filtering are given below. The media content filtering is done by performing skin tone detection. Total no of pixels can be identified from the extracted image. By performing skin tone detection we can identify the total no of skin pixel in the extracted image.

1)Extract humans from image
2)Run skin tone detection
3)Calculate the filter value as total number of skin pixels/total number of pixels
4)Compare the filter value with thresholds

The probability that the image being bad can be identified through this way.The suspicious profiles obtained by friend invitation graph analyser can be given to user behaviour analyser to find out the probability that the user being Sybil. Once identifying such profiles it will be given for further blocking.

**Sybil Detection:** Given a detection threshold $\delta f$, we consider a node $u$ as Sybil if its global acceptance rate $GAR < \delta f$ .The Sybil accounts are further given for blocking. Vote Credence algorithm takes the friend invitation graph $G$ and a set of trusted "seed users" as inputs, and outputs a set of active Sybil's that send many friend requests to real users.

### 3.4 Steps of Vote Credence Algorithm

Global Acceptance rate calculation
1. Initialize trust score ->0;
2. Get Initial ACE:
ACE=Total accept requests/ (Total no of accepted request +Total no of reject request);
3. Get message post rate:
MPR= Total no of successful message/ (Total no of successful message +Total no of failed message)
4. Calculate propagate trust score:
PTS =Count of no of friends having acceptance rate >0.5/Total no of friends.
5. Calculate global acceptance rate:
GAR= (ACR+MPR+PTS)/3;

**Text Filtering**
1. Split total message to obtain total no of words;
2. Identify Word, Word group in the message;
3. Increment count of identified word group;
4. Calculate the probability;
Probe = count of word group /total no of words;
5. Take the threshold of identified word group;
6. If Probability >Threshold
Block;
7. Else
Successfully send the message

**Media content filtering**
1. Extract human part from the image;
2. Image pixels = image. Width * image. Height;
3. Obtain cnt->total skin pixels
4. Color.B / color.G <1.249)
5. Channel Sum (color) / (3 * color.R) >0.696)
6. 0.3333 - color.B / Channel Sum (color) >0.014)
7. color .G / (3 * Channel Sum) <0.108)
8. fval = cnt / (image pixels)
9. If (fval >.25)
Block
10. Else
Successfully send

## 4. Conclusion

This paper presents Vote Credence, a system that uses both friend invitation graph and user behaviour analyser to defend against Sybil attacks. The existing method only considers Friend invitation graph analyser. So we introduce vote

credence based on both friend invitation graph and user behaviour analyser. First, we introduce friend invitation graph for Sybil defence, which nicely combine link structure and user feedback. Second, we propose new techniques including global vote aggregation to exploit the negative links.This improve the performance of this system.

## 5. Future Scope

Vote Credence, is a system that leverages user interactions of initiating and accepting links to defend against Sybil attacks. By implementing learning in the system, it is possible to establish a future extension to the current work. In current system if there is any bad content the Message will be filtered out. So learning can be including as a future extension in to the system which sends all the messages.

## References

[1] Lei Jin, Xuelian Long, Hassan Takabi, James B.D. Joshi \Sybil Attacks VS Identity Clone Attacks in Online Social Networks ","Year: 2015, Volume: PP, Issue: 99.
[2] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman,\Sybilguard:defending against sybil attacks via social networks,"in Proc. of SIGCOMM, 2006.
[3] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao,\Sybillimit: A near-optimal social network defence against sybil attacks,"in Proc. of IEEE SP, 2008
[4] G. Danezis and P. Mittal,\Sybilinfer: Detecting sybil nodes using social networks,"in Proc of NDSS 2009
[5] N. Tran, B. Min, J. Li, and L. Subramanian,\Sybil-resilient online content voting,"in Proc. of NSDI 2009.
[6] Viswanath, B., POST, A., Gummadi, K., And Mislove, A.\An analysis of social network-based sybil defenses."In SIGCOMM 2010
[7] N. Tran, J. Li, L. Subramanian, and S. S. Chow,\Optimal Sybil-resilient Node Admission Control."in INFOCOM 2011
[8] J. Xue, Z. Yang, X. Yang, X. Wang, L. Chen, and Y. Dai, “Votetrust: Leveraging friend invitation graph to defend against social network sybils,” in Proc. of INFOCOM, 2013.

## Author Profile

**Anagha P B** received the B Tech degrees in Computer Science and Engineering from Cochin University in 2013.She is currently persuading her M Tech in Computer Science and Engineering from University of Calicut.

**Janitha krishnan** received the B Tech and M Tech degrees in Computer Science and Engineering from university of Calicut and Anna University in 2008 and 2011, respectively. During June 2011-December 2012, she was working as assistant professor in Vimal Jyothi Engineering College. She is working with AWH Engineering College since December 2012.