TPFB: Trusted Path and Friend Behaviour with Thresholds Based Sybil Attack Detection for Social Network

(1) Vinod kumar Pandey  
M. Tech(CSE)  
Subharti Institute of Technolog &Engineering  
Swami Vivekanand University Meerut (U.P.), India-250005

(2) Sandeep kumar Patel  
M. Tech(CSE)  
Subharti Institute of Technolog &Engineering  
Swami Vivekanand University Meerut (U.P.), India-250005

(3) Mr.Saurabh Kumar  
Subharti Institute of Technolog &Engineering  
Swami Vivekanand University Meerut (U.P.), India-250005

Abstract- Community network is the group of nodes sharing the resources and communicates with each other using some relationship between the nodes. In such network the authenticity is measured using various security measures, but to judge the legitimacy of the node there is no such effective mechanism available. A Sybil attack involves losses related to fake or incorrect identities which somewhere in future forge the data and let the network performance degrades. Such attacks could be of fake profile types, or access request or some other user based service having incorrect information. If the entry is made incorrect at the time of registration then its detection gets failed. These fake identities needs to be removed by which devices and users gets assured about the systems safety and security. During the last few years, Sybil attack ratio is grown exponentially with increase in users and community based services, hence it notified triggering condition of criticalities of security. There are some of the existing approaches available to solve Sybil related issues, but still few remains unsolved. Most of them worked with trust based accurate Sybil profile detection with lower accuracy and detection rates which is not promising for futuristic growths and demands. This paper proposes a novel TPFB (Trust Path and Friend Behaviour) based Sybil attack detection and removal mechanism with real time outcomes. The approach uses trust certificates, threshold values, path rank values and continuous monitoring with historical data analysis for Sybil detection and removal. The trust is also forwarded by each node for very successful activity or constructive activity. The trusts also gets reduces for each destructive activity. At the point of evaluations, analytical verifications and judgments shows the effectiveness of the suggested approach over the existing ones.

Index Terms- Social Network, Sybil Attack, Behaviour Analysis, Trusted Path, Verification Certificates, TPFB (Trust Path and Friend Behaviour);

I. INTRODUCTION

Interest is fastest growing information flow network where the number of users counts ad their types are getting exponentially increased. Along with the user the number of nodes counts and the device dependent environment making the access more critical in terms of security. For controlling the changes performed by unauthorized users some authentication mechanism is required having more systematic verification parameters rather than only checking credentials.

As of now most of the web based applications are involved with cooperative communications which requires trust of one person or operation on other. Thus, the environment where the portability of users and applications are very high, few additional security mechanisms are required. The tradition security mechanism is only text based and will check some entered pattern of text only. But as of now with large user’s applicability areas, improved mechanism with continuous monitoring of processes are must for secure communications. Aim is to make the identity verification for the users accessing the service or data in an un-trusted environment. Assumption is made that each user needs to prove its identity with some process of trust factors calculations, if the value is high profile is taken as authentic and if the factor of relationship is weak then the users is treated as fake user with misleading profile.

Whenever the communication takes place between the two entities, packets gets travelled from various intermediate environments where the several users are trying to hold the data with some predefined hacking tools. Such communication must be making secure by using trust factor of each devices. If the devices or users are previously involved with any of the data affecting activity than its trust is dropped, and if the user’s activity is safe then the trust is increased. The device or user with lower values might cause data loss in near future. Now if in case the attacker is controlling any device as a legitimate user then the system is not capable of detecting such activities and consider the users as a actual and authorized user. This incorrect assumption makes the system down somewhere in future. Also the existing system is not able to detect the fake or incorrect profile created by some misleading or attacker user and by this activity the users gets access to the other trusted systems and users gets falsify.

Usually such critical consideration of security scenarios is taken over for distributed systems but as of now such technologies is common in other internet or web based applications automatically it spreads everywhere. Normally it is defined that for a system a user can control only single identity for services, but intentionally some users try to create some more profiles and fields which are fake but still control by the same user. Now by using these profiles such user can theft or mislead other users and causes their trust down for
further application usages. Thus detection of these misleading fake profiles for the system is termed as Sybil attack detection. Here the user perform identity theft and use them for several service acquisitions illegally. These bogus identities replaces the honest identities in a variety of tasks, including online content ranking, Byzantine failure defences [1], DHT routing, file sharing, reputation systems.

Various similar attacks are planted in several other domains also like sensor network, social network, community network, cognitive network, utility based applications etc. These attackers uses Sybil identities for giving the malicious opinions to the system or demands something by which process performance and feedbacks are degraded. Normally, they are applied for manipulating the trust of the system so as gets benefitted from this activity. Sybil defence mechanisms are based on such activity identification by trusted node analysis and activity ranking. Node with effective connectivity with trusted nodes is also having better trust than any other node [2]. Some of the specific application based study OSN (Online Social Network) are presented on the same topic given in [3].

II. BACKGROUND

In distributed environment, the applications and their threads are processed from different locations and devices. The number of these devices is very large with their user’s counts which usually applied in peer to peer network. Here the nodes and users identities are varied frequently due to their wide applicability and changes involved. In such network where the nodes and users behaviour are changing abruptly, forecasting the behaviour and analysis of current intentions are very complicated task. Also the users holding more than one identity is also work towards misleading the system. These identities are not duplicated but they have uniqueness so as to make fool of the system and cause misguided comments or incorrect service demands whatever be the intentions to make the actual working affected. These incorrect and fake identities are known as Sybil and the misleading actions of such identities are called as Sybil attacks. It could be bitterly understand by the Voting system example, here if the fake user with Sybil identify give the vote more than once with different identities than the actual results might be affected. The actual process is shown in figure 1.

Mechanism used to make the defence against those attacks is Sybil Detection mechanism and is dividing into two broad categories: Centralized and De-centralized.

First, the monitoring authority is responsible for detecting these misleading profiles with some central controlled authentication techniques. By this mechanism, the node cannot directly participate in distributed environment, if it gets registered then only the activity is performed. In second, the authentication responsibility is shifted from one to multiple nodes and systems. These are trust based systems which gets benefitted from this activity. Sybil defence mechanisms are based on such activity identification by trusted node analysis and activity ranking. Node with effective connectivity with trusted nodes is also having better trust than any other node [2]. Some of the specific application based study OSN (Online Social Network) are presented on the same topic given in [3].

II. BACKGROUND

In distributed environment, the applications and their threads are processed from different locations and devices. The number of these devices is very large with their user’s counts which usually applied in peer to peer network. Here the nodes and users identities are varied frequently due to their wide applicability and changes involved. In such network where the nodes and users behaviour are changing abruptly, forecasting the behaviour and analysis of current intentions are very complicated task. Also the users holding more than one identity is also work towards misleading the system. These identities are not duplicated but they have uniqueness so as to make fool of the system and cause misguided comments or incorrect service demands whatever be the intentions to make the actual working affected. These incorrect and fake identities are known as Sybil and the misleading actions of such identities are called as Sybil attacks. It could be bitterly understand by the Voting system example, here if the fake user with Sybil identify give the vote more than once with different identities than the actual results might be affected. The actual process is shown in figure 1.

Mechanism used to make the defence against those attacks is Sybil Detection mechanism and is dividing into two broad categories: Centralized and De-centralized.

First, the monitoring authority is responsible for detecting these misleading profiles with some central controlled authentication techniques. By this mechanism, the node cannot directly participate in distributed environment, if it gets registered then only the activity is performed. In second, the authentication responsibility is shifted from one to multiple nodes and systems. These are trust based systems which directs the misleading identities detecting and removal from their behaviour analysis. These decentralized process uses trust based mechanism for false profile identifications. These are further categorized into false positive and false negative attacks as explained by the paper [4]. There are various terminologies frequently used for any Sybil detection mechanism to describe the attacker’s behaviour are given: Insider vs. outsider; Selfish vs. malicious; Directed vs. in directed communications; simultaneously vs. gradually obtained Sybil identities; Busy vs. idle; Discarded or retained.

To overcome such Sybil attack occurrence in a network there are rules which needs to be reflected in execution environment by which effective and early identification of forged entries can be measured. These rules are given as:

(i) Secure routing: if an honest node X performs a lookup for an identifier ID, then the lookup mechanism must return the global successor of ID (present in the routing tables of honest nodes).

(ii) Pseudonymous communication: an adversary should not be able to determine the IP address corresponding to a user.

(iii) Privacy of user relationships: an adversary should not be able to infer a user’s social contacts.

(iv) Low control overhead: the control overhead of the system should be small to enable a scalable design. This excludes flooding-based and single-hop mechanisms.

(v) Low latency: the length of the path used to route to an arbitrary identifier should be small, in order to minimize lookup latency.
(vi) **Churn resilience:** even when a significant fraction of nodes fail simultaneously, lookup queries should still succeed.

(vii) **Fully decentralized design:** we target a fully decentralized architecture without any central points of trust/failure.

### III. LITERATURE SURVEY

Sybil attack is the most difficult to detect because of its behavioural elements and can be identified using pattern based activity analysis. As this, identification of fake profiles totally depends upon the other users, accessing or get in touch with these profiles can help in Sybil removal. Removal of such fake profiles can be achieved in timely manner before any major losses. During the recent few times various approaches related to effective Sybil detection is presented and be summarized as:

In the paper [6], the Sybil Guard is presented which will serve as detection mechanism using minimization of corruptive influence affects. The suggested protocol is based on community network for identities verifications using descriptive profiles for each user. The work gives a visual representation of trust relationship that might be created by mis-fishing nodes. This unknown node profiles needs to be detected on time by using some mechanism as proposed by the work. It also categorizes the actual profiles of legitimate users from the fake profiles from Sybil users. It enables a relationship that an honest node can accepts, and also is accepted by, most other honest nodes.

In the paper [7], suggested a learning based statistical model for Sybil detections in collaborative or community network. The model is known a latency community model (LC). The network use the partition boundary for uneven behaviour detection by which the fake or falsify entries is confirmed. The theoretical base of the paper confronted the attack category in dense use activities using Bayesian inference mechanism associated with some MCMC model. Evaluation of work shows experimentally that LC-based Sybil detector competes well with algorithms for the Sybil detection from the network security literature.

Sometimes these Sybil detection is gets more complex due to its execution environment like in distributed computing. Here the communities’ identities verification depends upon more than one user or machine which makes the task more complicated due to its dependencies. These can be further controlled by some contamination criteria based on trust evaluation. By this model the users verification for performing the particular task depends on the community values collectively verifies for computational loads. It also gives mechanisms to prevent the malicious influence of misbehaving nodes that do not perform the computational work [8]. The work also proves strong evidence that Sybil-Control can be practically deployed.

Sybil based vulnerability is detected in peer to peer systems with some modification using Sybil Limit tool proposed in [9, 10]. It exponentially improves the performance of existing Sybil Guard in near optimal time constraints. Some of the points where the tool provides benefits over the existing mechanism are given as: independent access of route optimization with random route values, works on intersection of vertices rather than focusing on nodes. The tool is also capable of balancing the load generated due to these fake profiles with effective benchmarking for safe estimates. Lastly, the outcome on real-world social networks inveterate their fast-mixing property and, thus, validated the elementary statement behind SybILimit’s approach.

Another effective tool is SybilShield [11] which uses multi-community model structure for malicious profile identifications. The work categorizes the actual behaviour profiles from the uncertain behaving users profiles by which the falsify nature is identified. SybilShield uses various agents for validating the random routes from where the user is accessing the data. If same route and approach is followed each time then the profile is of legitimate user and if some changes are made each time while pertaining the data then it indicates the malicious behaviour. Through the hypothetical probability examination and experiments on the MySpace data set, SybilShield is shown to greatly outperform existing mechanisms for reducing the false positive rate while trust the effectiveness of identifying Sybil nodes with an acceptable substitution.

Apart from the route based and trust based Sybil detection, some of the work suggests historical data analysis based behavior detection such as KD tree Sybil Identification [12]. The approach uses newly suggested algorithms SICT using connectivity threshold and another algorithm is SICTF using additional frequencies of data access. Both the algorithms are combined with previous Improved KD-Tree algorithm for community mining. Experimental evaluation shows that for measuring Sybil through striking event, the false positive is reduced than the existing random walk algorithm.

So many other specific approaches are available while considering situation based detection and saving the data losses such as: Privilege Attenuation [13], PrivacyJudge [14] and Trust in peer-to-peer systems [15]. Mainly it is been seen that most of the authors had worked with trust and threshold factor for accurate and timely detection. The work benefitted in various application domains such as used in banking sector for fraud transaction detection etc.

In the paper [16], STor is presented which is social network based anonymous communication detection using trusted router fro protected transmissions. It is a fuzzy based trust model to analyze the relationship between the friend’s profiles and the Sybil profiles using qualitative and quantitative attributes. The tools user selects routers by taking into account their trust in those routers. Here the work had also suggested algorithms for identification of friends community network and other trusted entities over the network. The experimental results show that STor can effectively establish trust-based circuits in a decentralized environment.

In many studies over Sybil some of the work had focused their intensions on deployment based detections and finding the countermeasures against them. The paper [17] shows that the world-wide deployed KAD network suffers large number of disbelieving insertions around shared contents and enumerate them. It detects the attack after analysing the sharing of the peers’ ID found around an entry after a DHT lookup. The work also evaluates the resolution and show that
it detects the most efficient configurations. It is able to detect the false-negative rate, and that the countermeasures successfully filter almost all the apprehensive peers.

IV. PROBLEM STATEMENT

One Sybil attack is mainly involved with malicious profiles whose intensions is towards making the process down or performs some uncertain data loss actions. As the profiles with incorrect entries cannot be defended by protecting their creations, thus the sorting of legitimate behaviour from malicious behaviour is performed. As the systems and the number of nodes gets expanded the vulnerability related to Sybil is also increased. Social and community networks are the major are on which the Sybil based forged profiles are available and to detect them is considered as a critical task. Existing algorithm used for such detection are well formed for small systems but for data intensive and computationally complex system the detection is quite difficult. The detection is depends on number of Sybil nodes counts and the surrounding environment by which these malicious activity gets started. The approaches limit the number of edges from which the attack is planted through a graph based mechanism. These scheme works towards the honesty measurement of each node and the respective edges. Some other process uses random walks within Sybil regions, and then the overall coverage can be measured. After studying the above mentioned approaches regarding the detection and removal of Sybil attacks or bogus identities here are the few problems which limits the capabilities of social networks are:

Problem 1: Effective Scalable Sybil attack detection for community network with effect analysis for each node and respective edges.

Problem 2: Tolerance approaches is not mentioned with existing schemes by which if a legitimate node some time behaves uncertain, then it could not be treated as Sybil node.

Problem 3: Existing approaches is totally relied on community nodes for such detection which will not gives accurate results every time. Thus some other mechanism needs to be embedded with those for further improved detections.

Problem 4: Decentralized approach is required for large peer to peer network and mobile networks.

Thus for Sybil defence schemes to work well, all non-Sybil nodes need to form a single community that is distinguishable from the group of Sybil nodes. So for improved and timely identification of such attacks fake user entry needs to be identified in accurate manner. Thus this work proposes a novel model for such detection.

V. PROPOSED TPFB APPROACH

This work proposes a novel Trusted Path and Friend Behaviour (TPFB) based accurate Sybil attack detection for community network. As per the suggested approach the work categorizes the actual legitimate profiles of user from the forged or fake profiles though some parameters that are trust dependent.

These parameters measure the trustiness of each node and edges using behavioural checks. The node with known and similar values previously seen can be taken as friend nodes and the nodes with newer access request form some novel path is taken for verification. According to Sybil detection, the aim is towards identification of fake or duplicate identities with same or different users having malicious and loyal behaviour. Such incorrect profile identities needs to be removed before any uncertain loss or theft conditions by which the systems actual performance gets degraded. The suggested approach is a combination of multistep procedure from data capturing to Sybil node detection and removal.

Let us consider the networked environment where nodes are communicating with each other by using various profiles. According to those profiles, these nodes are appears as a actual or legitimate node but planted to do some losses to the system. Now, in such network the nodes are frequently generates the data with respect to their hosted communications with several user identities. This generated data is later on analysed after the process of communication gets completed.

In cloud be taken as historical data analysis based Sybil detection. The aim is to identify the incorrect behaviour of the node. This historically generated data is used for identifying the nodes activities during the last few sessions. By this the loyalty and friend behaviour of node is detected from the other connected node in network. Here the behaviour verified with other node in the network. Now this collected data is passed to friend behaviour analysis module in which the trust with respect to each node is measured.

In the trust based friend behaviour analysis module, primary task is to detect the trusted identities with nodes from which these identities belong. If some node is connected to these trusted zones then the request coming from such elements is taken as trusted. And if the request is coming out from some other node not having trustiness, then it could be taken from Sybil identification process. Also this modules work for random path having maximum trusted intermediate node, and hence the belongings to these path is taken to be secure and legitimate. Here the trust certificate is generated for each node and path having a successful data transmission, and if the node is dropping the data or performing some uncertain activities.
So each node verifies a trust certificate of other each node in the same network and for accessing the controls from some other network they must contains some defined certificates whose count is more than a threshold. It also checks whether the given trust values and path ranking is greater than a decided threshold. If it is above then it proceeds further to communicate with its more friends or else communication is aborted.

The prime concern about such user and community management is for large networks which stores Tera Bytes of data and which has to be processed frequently to get the intelligent business decisions. Community networks spreads the information very fast so as to identify the bogus or fake profiles is very complicated task. This can be done only by confirming through some friend node. This node is called trusted behaviour identifier or trusted identity. From this identity path rank, behaviour of other users, certificates and thresholds can be easily identifiable from which accurate decision can be takes.

The above ranking and safe friend path identification improves the detection of fake identities by the behaviour analysis of each node. For each successful transmission the trust on the profiles gets increased. The aim is to make the device more intelligent so as to make the difference between the legitimate and the fake profiles. This can be done only by confirming through some friend node. This node is called trusted behaviour identifier or trusted identity. From this identity path rank, behaviour of other users, certificates and thresholds can be easily identifiable from which accurate decision can be taken.

Also the work is gets the network monitored regularly for several other Sybil based detection conditions measures as a pattern. These patterns could be of any type having predefined uncertain conditions. The node make the partition sin the network is also taken out from other node. On the behalf of these checks the authority node issues a trust certificate to other node. The node and route having maximum trust certificates will be taken as normal or legitimate route with proper behaving identities or profiles. But any irregularities assure some Sybil nature. The above process is the regular process and by which continuous monitoring of the network is performed.

After analysis of such Path Rank, Trust Certificates Verification, Threshold Condition and Regular Behaviour Monitoring Sybil node detection and removal can be accurately done. At the initial level of our work it seems to be giving better results than any other existing approaches. It can be implemented for so many domains like, Social Networks, VANET, MANET, and community network etc. Later versions of this approach will definitely give improvements in social media over mobile devices where verification of entities is quite difficult.

Expected Outcomes
1. It is useful to find Sybil Attacks thus making the system robust.
2. Easy detection using Path rank, Trust Certificate and Threshold values.
3. Trust certificates allows only authentic user to further forward the request.
4. Business intelligence applications more likely to be motivated after Sybil removal.
5. It is used to find fake user identities.
6. It is feasible to limit the number of attack edges in online social networks by relationship rating.

Application Domains
(i) Mobile Networks
(ii) Community Network
(iii) Auditing and Compliance
(iv) Cash Economies  
(v) Reputation Systems etc

VI. EVALUATIONS PARAMETER

Sybil attack detection improves the network performance and makes it more secure against some uncertain and forged process by some fake identities or profiles. As these attacks is mainly based of user provide entries, thus to detect the false entries is again a complicated task. During the last few times various approaches had been developed to overcome the issues related to the Sybil attack is given. This paper also proposes a novel mechanism to further improve the detection rates with minimum computation involved. Thus the goal of the work is to detect Sybil identities and reject them from further accessing the services using community based trust model. To access the performance of the work there needs some parameters for comparison purpose. These parameters are given as:

- **Robustness:** It is measured as effective protection of community network from Sybil attacks and up to which content this detection is accurate and on time. It is also measured by false positive detection rate in unit time.
- **Overhead:** This detection must generate some of the computational complexity for the system. So these complexities must be low as per the value of information is concerned. Hence it works for time, storage, and communication overhead.
- **Accuracy:** As the detection is totally based on trust computation thus if at some points of time the approach fails to detect the attacker than heavy losses are associated with the factors. Also missing or incomplete detection is similar to not detecting the Sybil, thus if it occurs the accuracy of the systems gets down. Thus accuracy is an important parameter.

VII. CONCLUSION

This paper presents a novel Sybil defence mechanism using TPFB (Trusted Path Friend Behaviours) based Sybil identities detection. The work uses existing topological analysis of each node and their respective activities of last few sessions for behaviours. Such measurement is a type of pattern by which futuristic entries is compared. Initially some nodes with actual behaviour will generate the data. Now this data is analysed using some trusted path and node values based on successful transmission and actual activities. If the node verifies this level then its entries are considered to be original, and if the node fails to overrule the threshold then it will be taken as Sybil profile. At the initial level of work, the proposed approach is capable of identifying the Sybil and in on time. Analytical evaluation will also show the effectiveness of the suggested approach.

FUTURE WORK

Forge identities detection is required some early decisions like verifications of user's intention at the time of profile creations. This can be achieved by historical transmission and activity details analysed in real time which demands heavy processing requirements. Thus in near future the practical implementation of suggested approach and its verifiability have to be done which later on extended in some real application.

REFERENCES