Survey of Literature Available on Sybil Defense Techniques

Suchita M. Bonde[1], Prof. Chinmay Bhatt [2]
Department of Computer Science & Engineering
[1,2] RKDF Institute of Science and Technology (RKDFIST), Bhopal (MP), India.

Abstract- *The problem of malicious activities in online social networks, such as Sybil attacks and malevolent use of fake identities, can severely affect the social activities in which users engage while online. For example, this problem can affect content publishing, creation of friendships, messaging, profile browsing, and commenting. Moreover, fake identities are often created to disseminate spam, use the private information of other users, commit fraud and so on. A malicious person can generate numerous fake accounts for these purposes to reach a large number of trustworthy users. Thus, these types of malicious accounts must be detected and deactivated as quickly as possible. However, this objective is challenging because a fake account can exhibit trustworthy behaviors and have a type of name that will prevent it from being detected by the security system. In this article, we provide a comprehensive survey of literature from 2010 to 2016 on Sybil attacks in online social networks and use of social networks as a tool to analyze and prevent these attack types. We first review existing Sybil attack definitions, including those in the context of online social networks.*

Keywords: sybil attack, Sybil Defender, Online social net-works

I. INTRODUCTION

As the popularity and influence of OSNs have increased, the incentives to attack these systems have also grown. A malicious user can create multiple false identities to gain access to sensitive private information to perform various types of several cybercrimes, such as compromising data integrity, trolling (making deliberately provocative or offensive online postings or criticism of opinions), rigging popularity, scamming, and breaking trust in online associations. These OSN attacks are types of Sybil attacks. The Sybil attack is named after the subject of the book Sybil, which describes a person diagnosed with dissociative identity disorder [10]. OSNs have become an integral part of contemporary life, with many people relying on them in the realm of work, social inter-actions, information sharing, and other aspects of daily living. Any negative impact on these areas due to Sybil attacks not only damages the user experience, but also the marketability and advertising potential of the given OSN. Outlined below are the key aspects that consider the above circumstances and motivate this work to survey defense schemes against Sybil attacks on OSNs.

OSN openness: OSNs are created as open platforms. To malicious users, the most attractive aspect of OSNs is their ready connection to many users at a minimal cost when compared to other internet channels. For example, in February 2010, the accounts of thousands of Twitter users, including the Press Complaints Commission and BBC correspondent Nick Higham, were hijacked after a viral phishing attack [1].

Social connection blind trust: Another major OSN loophole is that online users tend to trust their social connections and blindly value them. Thus, they may fall into a trap set by fake social connections perpetrated by cyber attackers. For instance, users are more likely to click a spamming link shared by a careless friend than the one they find on a random web page [2-3]. Spammers employ many techniques to send unwanted messages to users of OSNs, such as Facebook and Twitter. Such mes-sages or tweets typically present either advertisements or fraudulent information, thereby enabling the perpetration of phishing attacks or the spread of malware through embedded URLs. For example, in August 2009, it was reported that nearly 11% of all Twitter posts was spam [4].

OSN recommender system limitations: Recommender systems in OSNs are divided into four types according to the given OSN community: interest, friend, location and random-based communities [5]. The attackers strive to mimic the target profiles by building fake accounts that are consistent with one or more those of communities. Hence, the OSN recommender system directly chooses these fake accounts to recommend them to its users. Most OSN users prefer to have as many "friends" as possible to appear popular. Once the user adds the friends to her/his friendship list, the attacker can begin generating fake or otherwise malicious content [6-7].

Reputation system vulnerability: Similar to recommen-dation systems, reputation systems are likewise vulnera-ble to malicious users [7]. A reputation system calculates the numerical reputation of individual identities based on pairwise feedback between the identities [8]. A

significant need exists to prevent fake identities from artificially boosting the attackers reputation.

Fake ratings and reviews: Online reviews are helpful to online users and off and online buyers. Thus, online reviews are very popular in e-commerce systems such as Amazon, Yelp and Google Play. Recently, online companies have had to curb the rising number of fake ratings and reviews that promote personal interests or bolster undeserving products, restaurant, apps, etc. In simple terms, these fake ratings and reviews generally subvert the reputation of a certain product or service.

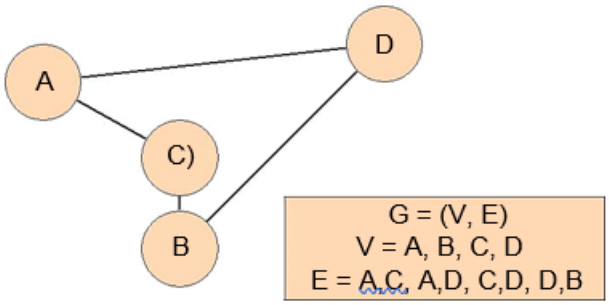

Fig. 1: Social network structure

The above aspects have made social networks attractive to Sybil attackers who manipulate large sets of malicious ac-counts to launch attacks [9-11]. Therefore, it is very important to detect Sybil attackers in an OSN to prevent their malicious activities.

The remaining section of the paper is organized as follows: In section II describe the Background overview of Sybil attack and related work about detection and prevention mechanism. In Section III describe Sybil Attack solutions and Schemes. Section IV section gives overall conclusion of the paper.

II. BACKGROUND

In this section, we provide a general background and overview of OSN and some of its properties that help de-fending against Sybil attacks.

1) OSN overview: An OSN can be described as a website that provides a venue for users to connect with other users, friends and family members [12]. An OSN user is represented by an account profile. These three termsuser, account and profile can be used interchangeably. The profile describes the users social attributes such as name, gender, interests, and contact information. The relationship between accounts can be either two-sided such as friendships in Facebook or one-sided such as a followership in Twitter [3]. Apart from creating reliable connections an OSN can enable a user to share photos, music, videos and other individualized information with certain friends or the general public [12]. In addition to Facebook, Twitter, YouTube some highly popular OSNs include LinkedIn, Google+, Instagram and Myspace. These networks are a very suitable way for enabling users to remain in contact with other users around the world. Moreover, they are convenient because they enable users to create new connections with people who have similar affinities, such as a similar profession or similar interests [13-14].

Diameter and mean path length: The diameter of a social graph is defined as the greatest distance between any two nodes. It can be described as the maximum value of the shortest path length between a pair of nodes in the social graph. For instance, if l(i;j) is considered to be the length between nodes i and j, then diameter d is considered the maximum l(i;j) of all possible node pairs. On the other hand, the mean path length lm, of a social graph is considered the mean distance between all the nodes that are located in the graph [35].

$$l_m = \frac{1}{n(n\ 1)} X_{l(i;j)} \qquad (1)$$

Power and centrality: Power in this context is defined as a basic property of the social structures that have close relations with centrality [14-15]. Various techniques have been used to facilitate research on the graph property of power. However, the three main measures used to describe power or centrality is the degree, closeness and betweenness. In this regard the degree refers to the number of edges of a certain node. This property is often normalized by the sum of the edges that are available in a graph. betweenness refers to the possibility that any node will require a different node in order to reach an additional node using the shortest path [14]. Betweenness is computed by using Equation 2.

$$B_i(v) = \sum_{i6=\ 6=j2v} \frac{p(V)}{p} \qquad (2)$$

where p is the total number of the shortest paths from i to j, and p(v) is the number of paths that cross the node.

Degree distribution: The degree distribution implies the distribution of degrees for all the nodes that are available in a graph. This property very closely follows a power law for all real social networks. A large percentage of nodes feature a minimal degree. However, a small number of nodes have a very large degree [16]. Power law is represented in Equation 3.

P (K) k (3)

where is a constant. In sum, social networks are im-portant sources of information, and they can be modeled using graph theory. However, they can be difficult to ana-lyze. The analysis may require considerable knowledge of a means to calculate the metrics and decipher appropriate meaning from the statistics obtained.

2) Definition of a Sybil attack: In OSNs a Sybil is a fake account with which a user attempts to create multiple identities to make as many friends as possible with legitimate accounts. A Sybil account can lead to many malicious activities in an OSN. For example, a Sybil can outvote legitimate users through Internet options [24]. In addition, it can control some accounts by using the fake identities to provide misleading information [16]. Moreover, false reputation can be created based on Sybil accounts [18]. Generally, a Sybil attacker is a user that creates many fake identities to increase the power and influence of the attacker in the network and thereby generate engage in malevolent activities, such as disseminating social spam, distributing malware, distorting online ratings, executing phishing attacks, and so on [19-20].

III. SYBIL ATTACK SOLUTIONS AND SCHEMES

In this section, we provide an overview of different Sybil schemes that have been used by researchers to execute various Sybil detection and prevention algorithms and tools. In gen-eral, the Sybil schemes are divided into four main categories:

Graph-based schemes, machine-learning-based schemes, man-ual verification and Prevention approaches. Graph-based (also called network-based [20]) schemes are further divided into subcategories: Sybil detection and Sybil tolerance. Meanwhile, the machine learning schemes are divided into respective supervised, unsupervised, and semi-supervised methods. The classification approach is based on our perspectives and un-derstanding of the given problem.

A. Graph-based methods

Graph-based methods use social network information in the case of Sybil to represent inter dependencies between objects using edges or links. These methods strongly rely on social graph properties to distinguish Sybil users from legitimate users. They can be classified into Sybil detection and Sybil tolerance schemes, respectively. Most of the graph-based solutions rely on the assumptions outlined below.

1) Sybil detection schemes: Most of the Sybil detection schemes are based on the concepts of the graph random walk and mixing time. Some of the most common approaches are social-graph-based techniques, such as the first ones proposed for this purposeSybilGuard [10] and SybilLimit [22]among other Sybil-defense approaches. Other approaches have in-cluded Sybil-resilient systems that use application knowledge to control Sybils. Below all of these schemes and methods are detailed.

SybilLimit: In addition, similar to SybilGuard, SybilLimit relies on an estimation procedure when determining the length of the random walk as well as the quantity of random walks that are required. In general, these two parameters are crucial for determining the occurrence of a cutoff.

SybilShield: is intended to address the limitations of SybilGuard, whereby a legitimate suspect node can be treated as a Sybil node. It achieves this objective through modified random walks, known as the random route approach, whereby a suspect node is deemed legitimate if the random routes of a suspect node and verifier intersect. The main concept is to provide an opportunity to suspect nodes through an agent walk approach to mitigate the problem. Additionally, agents are selected from communities other than the verifiers community.

SybilDefender: The general design of a SybilDefender scheme includes a Sybil identification algorithm that identifies the Sybil nodes, a detection algorithm for the Sybil community, and two approaches that limit the quantity of attack edges experienced in the OSN. A major drawback of the SybilDefender scheme is that it only relies on performing a limited number of random walks within the social graph.

SybilRank: it influences its support for various trust seeds to reduce the number of false-positives that result from the existence of multiple communities considered non-Sybil. Secondly, the scheme enables a very flexible seed selection process that makes it much more difficult for attackers to target these seeds. Furthermore, its effectiveness is only moderately decreased when the distance of the Sybil decreases in relation to that of the trust seeds. This scheme is thus suitable for large-scale attacks in which fake accounts can be developed and maintained at an extremely low cost. In addition, SybilRank can be deployed on a social graph that features strong relationship edges.

Integro: It is worth mentioning that Integro is limited to undirected social graphs and it delays the consideration of new accounts. Thus, it has a major limitation in terms of the accuracy of Sybil detection among new users. It is furthermore

limited in its method of correctly detecting Sybils in cases in which a large number of fake accounts can collect more trust than a low number of real ones.

2) Sybil tolerance: Sybil tolerance methods limit the effects of Sybils that are present in an OSN. These defense approaches utilize many techniques for providing a particular credit for each edge in the graph. They ultimately limit the impact of attack edges. The networks that apply this concept are called credit networks, wherein a node trusts another node by giving a pairwise credit to its link up to a certain limit. The three renowned approaches in this classOstra, SumUp, and TruTopare detailed below.

Ostra: Ostra requires that recipients classify incoming communication as either wanted or unwanted. Providing explicit feedback is a slight burden on the user. Ostra can be used only in conjunction with the invitation-only social network. The authors were unable to obtain a communication trace of the same scale as the OSN that they considered. Consequently, they made assumptions about the likely communication patterns in the OSN. They concluded that users communicate with nearby users much more often than with users who are far away in the OSN.

SumUp: This scheme uses Cmax in deciding the max-imum quantity of votes that should be accepted by the system. SumUp can prohibit an attacker who progres-sively misbehaves. This scheme additionally assumes that a mini-cut exists between legitimate nodes and a cote collector occurs at the collectors. However, between the Sybils and legitimate nodes, the mini-cut occurs at the attack edges. Nevertheless, SumUp accepts 0(logn) Sybils or attack edges, and it requires knowledge of the overall system. Major disadvantages of SumUp are its high respective computation and run-time requirements.

canal: algorithm which routes credit payments through landmark nodes. Canal consists of two components: 1) Universe Creator Processes-It Continuously selects new landmarks 2) Path Stitcher Processes-It continuously processes incoming credit payment requests. Canal uses these components to continually calculate new landmarks in parallel with doing flow calculations.

TrueTop: Several design choices for TrueTop repre-sent the incomplete evaluations involved in the research methodology. Thus, the likelihood was shown that various possible strategies can be used by attackers to gain control of OSNs. In addition, the findings provided only a few relevant results and showed that the other design choices and attack strategies would yield similar results.

B. Machine-learning methods

In this section, we provide an overview of the different supervised approaches that have been used by researchers to detect Sybil attacks in OSNs. The advantages and drawbacks of each method are additionally highlighted. Machine learning is a technique for autonomously acquiring and integrating knowledge obtained from experience, analytical observations, etc. [25]. It is usually divided into two main types: Supervised learning, such as regression models, naive Bayes, support vector machine (SVM), and decision tree models. Unsuper-vised learning, such as clustering algorithms (K-means, fuzzy C-means, hidden Markov models, etc.). Machine learning techniques are designed to solve problems involving massive amounts of data with many variables. These techniques are commonly used in areas such as pattern recognition (speech and image processing) and financial algorithms (credit scoring and algorithmic trading) [26].

1) Supervised methods:: Currently, some of the most commonly used supervised detection methods are the naive Bayes, SVM, nearest neighbor (e.g., k-NN), linear/logistic regression, and least squares models. In sum, the success of supervised methods is highly dependent on the use of domain knowledge of data to construct features. This concept is also known as feature engineering [27]. This approach can be time consuming compared to feature selection, which only involves returning a subset of important features. It can be effectively illustrated by the logistic regression models in OpinSpam, which had to be fed 36 features in the purview of content and behavior information. Experts with extensive knowledge of Amazon and its corresponding review dataset were therefore required. Few studies have been conducted to detect Sybil attacks using a supervised technique. First, OSN providers do not give access to their databases; access is given only through public APIs. Thus, building the ground truth is the default. Second, training data on a large scale is difficult to achieve without full access. Despite those obstacles, some researcher have obtained access to RenRen [28], the largest Facebook-like OSN in China. They performed a supervised technique to distinguish between Sybil and non-Sybil accounts. That researched is outlined below.

Uncovering Sybil [21]: The main aim of this research was to make two key contributions to Sybil detection in OSNs. The first was to use available data on the charac-teristics of Sybil in the wild to improve a feature-based Sybil detector that was adopted in RenRen. The author first showed how a properly crafted detector could find 99% of Sybils with very low false-negative and positive rates. The second contribution was to characterize the topology of Sybil graphs in OSNs, marking the first topology characterization of its kind. The author showed

that Sybils do not abide by the assumptions previously used in detectors behind community-based Sybil detec-tion algorithms, e.g., SybilGuard, SybilLimit, SumUp, SybilInfer, etc. Thus, the author closely examined if the Sybil accounts had some form of relationship in the background. The results showed that Sybil accounts actually colluded to execute attacks. The author then studied new forms of attack models that could be used to bypass the detector. It was shown that closely interwoven communities between Sybil users could succeed in this task. Using this information, the author demonstrated that the addition of a parameter, known as the external acceptance ratio, could improve the efficiency in Sybil detection as new attack models emerged. In addition, the study compared the performances of two Sybil detection algorithms: a simple threshold detector and a complex learning algorithm detector, SVM. The performance of the tuned threshold detector was virtually equal to that of the rather expensive SVM. The study also showed that use of the detector could be inaccurate when Sybil attacks employ new model-forming tightly knit communities. A better detector was determined to be one that combines a feature-based detector and an enhanced community detection approach.

Social Turing Tests [29]: The large study presented in [30] explored the feasibility of humans detecting Sybils in OSNs, while analyzing the detection accuracy of turkers. Under a certain set of conditions, the experts found that turkers delivered optimal results, while humans exhibited no uniform detection accuracies. The author analyzed this data and used it to drive the design of a multi-tier crowdsourcing system. The study of the system showed that it could be quite efficient as a standalone detection system or as a complimentary system to other detec-tion techniques. The study of the human crowdsourcing capability was performed with the aid of a corpus of ground truth data Sybil accounts accrued from RenRen and Facebook. The two datasets had three sets of profiles: confirmed Sybils, confirmed legitimate users, and sus-pected profiles. Using crowdsourcing was not appropriate, however, because most members of the crowd lacked relevant expertise, especially in terms of being able to verify turkers. As a consequence, crowdsourcing was found to be inherently inconsistent in its assessments [31].

2) Unsupervised methods: Some researchers have focused on detecting Sybil attacks in social media [32] using clustering and latent variables techniques [33]. Clustering is the process by which objects with similar characteristics are grouped in a set. Scavenger, for example, looks for URLs in Facebook wall posts; the URLS are then used to build a wall-post similarity graph. The graph is then used to select clusters that are most likely to result from spam campaigns. A significant feature of clustering is that it can uncover even hidden structures in unlabeled data while also having the potential to summarize key features [34]. Latent variable models are other clustering models that employ latent variables as a representation of hidden variables. Examples include spamicity of reviewers in Amazon. Spamicity, in this case, is the likelihood of something being spam. These concepts are well incorporated in an unsu-pervised Bayesian inference framework known as the author spamicity model (ASM). The framework creates a hypothesis that differentiates fake users and from real ones in a behavioral perspective. This task is performed by representing these two clusters in the latent space. Evaluation results showed that unsupervised spamicity models are actually very effective [35].

Clickstream [35]: The authors presented a method that leverages clickstream models on the basis of experimental approaches. They analyzed clickstream activity of click patterns of normal and Sybil users by employing logs from different social networks. Clickstream models were then proposed to characterize the click behaviors of users. These models were used to create a Sybil detection system independent of input from a service provider. The unsupervised Sybil detector created by the author of [36] exhibited excellent performance metrics, especially in terms of accuracy in real-world OSNs, namely RenRen and LinkedIn. Of one million random user accounts fed into the detector, it detected 22,000 Sybil users. In LinkedIn, out of the 40,000 analyzed users, the detector identified approximately 1,700 as Sybils out of 4,000 that had already been identified as Sybils.

SybilExposer [37]:Authors proposes a different algo-rithm, SybilExposer, which deals with the weaknesses of manual inspection, high computation costs, and the need to consider various Sybil communities. This al-gorithm also enhances identification, and has over 20% capability to detect Sybil within a community compared to SybilRank. This algorithm can be used as a scalable first line of identification of various Sybil communities in very large networks. The simulation results indicates that the proposed algorithm performs much better in terms of computational costs and effectiveness, compared to other algorithms. However, future research directions should consider extending the scheme of the algorithm to deal with non-community Sybil nodes as well as edges that occur between Sybil communities.

3) Semi-supervised Methods: The semi-supervised learning is a technique that uses a set of labeled and unlabeled data. It is thus between supervised learning techniques that use labeled data and unsupervised learning techniques which use only non-labeled data. It has been shown that the use of unlabeled data, in combination with labeled data, significantly improves the

quality of learning. Many researchers employ this method for their works.

SybilBelief [23]: SybilBelief is a semi-supervised learn-ing framework that relies on loopy belief propagation and Markov random fields. The inputs of this scheme include the social network of the nodes in the system, a small number of known Sybils, and a small number of known benign nodes. In its operation, the scheme operates by propagating the label information from the benign node, which is known to the remaining nodes that are unknown in the system. In addition, SybilBelief can accurately identify the Sybil nodes with few cases of false-positive and false-negative rates. Moreover, it is resilient to noise and can perform orders of a magnitude better than other Sybil classification mechanisms. Experimental results ob-tained in [24] showed that SybilBelief performed an order of magnitude better compared to SybilInfer and SybilLimit. A major limitation with that study, however, is that the authors did not evaluate their approach using real Sybil users. In addition, it showed a lower accuracy when the number of attack edges increased, and it failed to work on weak trusted networks [65].

SybilFrame [24]: SybilFrame is used to solve the prob-lem of Sybil attacks in OSNs with weak trust. It enables the defending of this kind of network, while operating in conditions where the number of attack edges is large. It uses a multi-stage classification mechanism that analyzes heterogeneous sources and types of information on the profiles in the OSN. The authors of this method used data on suspended accounts to serve as the basic truth for Sybil attacks. This method can additionally be used to rank accounts in terms of trustworthiness. SybilFrame gathers all the information on the users to define if they are Sybil or not. During this stage, the computing of the node prior and edge prior information takes place. To conclude on edge priors, the authors assigned lower scores to attack edges and higher ones to the edges between the trustworthy nodes. Secondly, it correlates the information from local classifiers with the global properties of the network. In this stage, pairwise Markov random field and loopy belief propagation methods are used to make the conclusions on the posterior information. In the method framework, local classifiers are considered more effec-tive when combined with the global information. These authors also studied the structural differences of Sybil and non-Sybil nodes. They determined that most Sybil accounts are isolated, which makes it difficult to detect them, and new effective methods should be created.

C. Manual verification and user feedback

OSNs are targets for disseminating false and rumor-related content. This content proliferates every day and misleads thousands of OSN users. The content may include false news, doctored images, fake accounts, or any type of content that can be circulated among OSNs. Content can be analyzed in different ways to determine whether it is true or false, and it is the responsibility of the user to analyze content before commenting on or sharing it. Many social networks, such as Facebook and Twitter, enable users to report content that violates their rules or terms of privacy. This content could be spam, harmful or malware links, multiple accounts, unrelated content, or aggressive or abusive content. OSNs also encourage users to recommend any additional action to improve their security and privacy. Many OSNs have authenticated public figures and key brand identities and provided them with verified accounts. These accounts are distinguished in Twitter and Facebook by a blue badge, which can be a sign to differentiate fake or normal accounts from the verified one.

Typically, verified accounts are provided only to celebrities, politicians, and major brands; however, it is not impossible for ordinary people or small brands to obtain a verified account. Posts and contents of these verified accounts are most likely to be true.

D. Prevention Approaches

Traditional approaches to defeating Sybil attacks are based either on trusting central authorities or connecting identities to resources that cannot be easily obtained by the attacker. This prevents the attacker from creating Sybil identities at the outset. These traditional approaches are called Sybil preven-tion. Many OSNs adapt these approaches, such as Cyworld2, to prevent Sybil attacks. In these approaches, the users must provide verified identities when creating new accounts, such as a social security or passport number. A simpler approach is to enforce users to solve a challenge-response test, such as CAPTCHA or crypto-puzzles, as a requirement to obtain access to system services [91, 92]. A more recent approach is to verify the user identity by sending a verification SMS message to the users phone. Although this approach is widely used, attackers can bypass it by using virtual mobile number services, or by obtaining many accounts using disposable phones.

E. Other Supplementary Work:

Other researchers have moved to work on improvements on certain dimensions. For example, in [94] authors are trying to solve the problem of measuring Sybil groups that exist in large-scale online social networks based on

different levels. These levels include malicious activities, individual informa-tion, and social relationships among other related aspects. The authors try to leverage the characteristics of Sybil groups in order to improve the security mechanisms of OSNs to provide defense against Sybil attacks. The work in [37] try to solve the problem of retroactively identifying fake profiles by analyzing 62 million Twitter user profiles that are available in the public domain. The authors identify a sub-set of fake user accounts that are considered to be highly reliable after using a pattern-matching algorithm on the screen-names, as well as an elaborate analysis of the tweet update times. The authors try to reveal the distinct behavior of the user accounts that are labelled as fake, based on a ground truth data set, after analyzing the statistics of profile creation as well as the URLs of the accounts considered to be fake. Similarly, authors in [96] trying to describe an accessible approach that can facilitate the finding of groups that consist of fake accounts that are registered by a similar actor. The major technique that they employ in their study is the use of a supervised machine learning pipeline that is useful in classifying a whole cluster of accounts as either legitimate or malicious. In their research study, they make use of major features like name, company, email address, or university, and these details feature some essence of patterns within a cluster. The authors apply their framework during the study and analyze the account data collected on LinkedIn, as grouped by two factors: registration IP address, and the registration date.

Recent work by [36] looked to extend the work against Sybil attacks in distributed systems by creating a framework that accounts for the adversarys ability to create periphery attacks. This work was inspired by the shortcomings of graph-based Sybil detection models designed in prior literature whereby it is assumed that the number of links that an adversary can make between Sybil and honest nodes is restricted. While models leveraging this edge limiting assumption have worked against certain types of Sybil attacks, it has also been shown that they fail when Sybil attacks are characterized by isolated Sybils connected to many edges to honest nodes. This sort of attack is known as periphery attacks and they violate the edge limiting assumption thus rendering conductance-based Sybil defense strategies futile.

IV. CONCLUSION

In this paper, we surveyed state-of-the-art research relating to Sybil attack defense schemes and techniques in OSNs. We quantified these works from different perspectives, including the methodologies, algorithms, assumptions, and designed models. We further provided a discussion to summarize our observations based on the reviewed literature. This discussion may be useful for readers in gaining a better understanding of the problem of Sybil attacks and the methodologies that have been proposed to address them. Although various existing solutions are intended to resolve the problem of Sybil attacks in OSNs, they remain an immense problem. Further research is needed to stop fake users from causing negative effects, altering actions, and driving quantifiable outcomes while using online platforms.

V. ACKNOWLEDGMENT

We thank Mr. Yogesh Sanjiv Chaudhari for his support while doing this work and comments that greatly improved the manuscript. We would also like to show our gratitude to the Prof. Varsha Namdeo Head Computer Science Engineering Department from RKDF Institute of Science and Technology (RKDFIST), Bhopal (MP), India. for sharing their pearls of wisdom with us during the course of this research. We would also like to thank our family members for their love and support.

REFERENCES

[1] C. Arthur (2010). Twitter Phishing Hack Hits BBC, Guardian and Cabinet Minister. Available: http://www.theguardian.com/technology/2010/feb/26/twitter-hack-spread-phishing.

[2] L. Jin, Y. Chen, T. Wang, P. Hui, and A. V. Vasilakos, ”Understanding User Behavior in Online Social Networks: A Survey,” Communications Magazine, IEEE, vol. 51, 2013, pp. 144-150.

[3] M. Wani, M. A. Alrubaian and M. Abulaish, ”A User-centric Feature Identification and Modeling Approach to Infer Social Ties in OSNs,” in Proceedings of International Conference on Information Integration and Web-based Applications & Services, 2013, p. 107.

[4] SciTechBlog (2010). A New Look at Spam by the Numbers. Available: http://scitech.blogs.cnn.com/2010/03/26/a-new-look-at-spam-by-thenumbers/.

[5] X. Han, L. Wang, R. Farahbakhsh, . Cuevas, R. Cuevas, N. Crespi, and L. He, ”CSD: A Multi-User Similarity Metric for Community Recommen-dation in Online Social Networks,” Expert Systems with Applications, 2016.

[6] J. Tang, X. Hu, H. Gao, and H. Liu, ”Exploiting Local and Global Social Context for Recommendation.”

[7] H. Yu, C. Shi, M. Kaminsky, P. B. Gibbons, and F. Xiao, ”Dsybil: Optimal Sybil-Resistance for Recommendation Systems,” in Security and Privacy, 2009 30th IEEE Symposium on, 2009, pp. 283-298.

[8] M. Sirivianos, K. Kim, and X. Yang, ”FaceTrust: Assessing the Credibility of Online Personas via Social Networks,” IACR Cryptology ePrint Archive, vol. 2009, p. 152, 2009.

[9] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, ”SybilLimit: A Near-Optimal Social Network Defense Against Sybil Attacks,” IEEE/ACM Transactions on Networking, vol. 18, p. 885, 2010.

[10] H. Yu, M. Kaminsky, P. B. Gibbons, and A. D. Flaxman, ”Sybilguard: Defending Against Sybil Attacks via Social Networks,” Networking, IEEE/ACM Transactions on, vol. 16, 2008, pp. 576-589.

[11] A. ALGhamidi, ”General Investigations Shake Hands with the Media in a Joint Mission to Enlighten,” in Alriyadh News Paper, vol. 17048, ed. Riyadh, KSA: Alyamamh Press, 2015.

[12] R. Gunturu, ”Survey of Sybil Attacks in Social Networks,” arXiv preprint arXiv:1504.05522, 2015.

[13] S. Y. Bhat and M. Abulaish, ”Analysis and Mining of Online Social Networks: Emerging Trends and Challenges,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, 2013, pp. 408-444.

[14] A. Chinchore, F. Jiang, and G. Xu, ”Intelligent Sybil Attack Detection on Abnormal Connectivity Behavior in Mobile Social Networks,” in Knowl-edge Management in Organizations, L. Uden et al., eds., Heidelberg: Springer, 2015, pp. 602-617.

[15] M. O. Jackson, Social and Economic Networks vol. 3, 2008, Princeton, NJ: Princeton University Press.

[16] I. Stanton and A. Pinar, ”Constructing and Sampling Graphs with a Pre-scribed Joint Degree Distribution,” Journal of Experimental Algorithmics (JEA), vol. 17, 2012, pp. 3-5.

[17] M. Conti, R. Poovendran, and M. Secchiero, ”Fakebook: Detecting Fake Profiles in On-line Social Networks,” in Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), 2012, pp. 1071-1078.

[18] M. Y. Kharaji and F. S. Rizi, ”An IAC Approach for Detecting Profile Cloning in Online Social Networks,” arXiv preprint arXiv:1403.2006, 2014.

[19] K. Hoffman, D. Zage, and C. Nita-Rotaru, ”A Survey of Attack and Defense Techniques for Reputation Systems,” ACM Computing Surveys (CSUR), vol. 42, 2009, p. 1.

[20] H. Yu, ”Sybil Defenses via Social Networks: A Tutorial and Survey,” ACM SIGACT News, vol. 42, 2011, pp. 80-101.

[21] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove, ”An Analysis of Social Network-based Sybil Defenses,” ACM SIGCOMM Computer Communication Review, vol. 41, 2011, pp. 363-374.

[22] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai, ”Uncovering Social Network Sybils in the Wild,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 8, 2014, p. 2.

[23] N. Z. Gong, M. Frank, and P. Mittal, ”Sybilbelief: A Semi-supervised Learning Approach for Structure-based Sybil Detection,” Information Forensics and Security, IEEE Transactions on, vol. 9, 2014, pp. 976-987.

[24] P. Gao, N. Z. Gong, S. Kulkarni, K. Thomas, and P. Mittal, ”Sybilframe: A Defense-in-Depth Framework for Structure-based Sybil Detection,” arXiv preprint arXiv:1503.02985, 2015.

[25] D. S. Sisodia and S. Verma, ”Analysis of Spamming Threats and Some Possible Solutions for Online Social Networking Sites (OSNS),” Analysis, vol. 1, 2015, p. 27207.

[26] B. Batrinca and P. C. Treleaven, ”Social Media Analytics: A Survey of Techniques, Tools and Platforms,” AI and SOCIETY, vol. 30, 2015, pp. 89-116.

[27] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, ”Feature Engineering Strategies for Credit Card Fraud Detection,” Expert Systems with Applications, 2016.

[28] G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, and B. Y. Zhao, ”Social Turing Tests: Crowdsourcing Sybil Detection,” arXiv preprint arXiv:1205.3856, 2012.

[29] Z. Cai and C. Jermaine, ”The Latent Community Model for Detecting Sybil Attacks in Social Networks,” 2012.

[30] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, ”Detecting and Characterizing Social Spam Campaigns,” in Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement, 2010, pp. 35-47.

[31] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, ”Spotting Opinion Spammers using Behavioral Footprints,” in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 632-640.

[32] G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao, ”You are How You Click: Clickstream Analysis for Sybil Detection,” in Proc. USENIX Security, 2013, pp. 1-15.

[33] S. Misra, A. S. M. Tayeen and W. Xu, ”SybilExposer: An effec-tive scheme to detect Sybil communities in online social networks,” 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-6. doi: 10.1109/ICC.2016.7511603

[34] F. Li, P. Mittal, M. Caesar, and N. Borisov, ”SybilControl: Practical Sybil Defense with Computational Puzzles,” In

Proceedings of the Sev-enth ACM Workshop on Scalable Trusted Computing, Raleigh, North Carolina, USA, 2012.

[35] S. Pise and R. Kumar, "Recent Trends in Sybil Attacks and Defense Techniques in Social Networks," in International Journal of Engineering Research and Technology, 2014.

[36] Jiang, Jing, Zi-Fei Shan, Xiao Wang, Li Zhang, and Ya-Fei Dai. "Understanding Sybil Groups in the Wild." Journal of Computer Science and Technology 30, no. 6 (2015): 1344-1357.

[37] Gurajala, Supraja, Joshua S. White, Brian Hudson, and Jeanna N. Matthews. "Fake Twitter accounts: profile characteristics obtained using an activity-based pattern detection approach." In Proceedings of the 2015 International Conference on Social Media & Society, p. 9. ACM, 2015.