An Enhanced Sybil Guard to Detect Bots in Online Social Networks

Nisha P. Shetty, Balachandra Muniyal*, Arshia Anand and Sushant Kumar

Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal-567104, India
E-mail: bala.chandra@manipal.edu
**Corresponding Author*

Received 31 July 2021; Accepted 21 October 2021;
Publication 16 November 2021

Abstract

Sybil accounts are swelling in popular social networking sites such as Twitter, Facebook etc. owing to cheap subscription and easy access to large masses. A malicious person creates multiple fake identities to outreach and outgrow his network. People blindly trust their online connections and fall into trap set up by these fake perpetrators. Sybil nodes exploit OSN's ready-made connectivity to spread fake news, spamming, influencing polls, recommendations and advertisements, masquerading to get critical information, launching phishing attacks etc. Such accounts are surging in wide scale and so it has become very vital to effectively detect such nodes. In this research a new classifier (combination of Sybil Guard, Twitter engagement rate and Profile statistics analyser) is developed to combat such Sybil nodes. The proposed classifier overcomes the limitations of structure based, machine learning based and behaviour-based classifiers and is proven to be more accurate and robust than the base Sybil guard algorithm.

Keywords: Sybil guard; Sybil; random walks (RW); loop belief propagation; machine learning; behavior based detection.

Journal of Cyber Security and Mobility, Vol. 11_1, 105–126.
doi: 10.13052/jcsm2245-1439.1115

1 Introduction

Online Social Network (OSN) has become the new age favorite for individuals to meet new people, gather and disseminate information, create influences and so on. Such is the influence of these sites that in recent statistics [1] Twitter, Facebook, Instagram etc. rank both in top ten downloaded apps or frequently visited sites. Simplicity, with no cost account creation and usage, has attracted masses in huge numbers towards these sites. This open architecture of OSNs have attracted many frauds who malign the readily and easily available platforms to connect to people with the intention of compromising data integrity, trolling, rigging popularity, scamming, breaking trust in online associations etc. [2]. There are many notable stories where the fake news spread by such bots have influenced stock markets, manipulated election results etc. Bogus accounts of influential celebrities are created to endorse many products. Such attacks can be further escalated to more deadly phishing, social engineering and DOS attacks. Named after the case study of a woman with multiple personality disorder [3], a Sybil attack is a type of security threat when a node in a network claims multiple identities [4].

1.1 Examples of Sybil Attack

Voting System: In an online E-voting system an account gets to vote only once. Antagonists rig the system by creating multiple fake accounts and thus superseding honest user's decision.

Fake news spreading: Parody Twitter account of popular news site Times Now spread fake news in 2018, many false profiles and pages were created to influence US elections, recently fake accounts of India Pak officers were found spreading hate and falsified information. One of the most notable incidents which took place in recent times was the hacking of American Press pages by Syrian Electronic Army to spread fake news that the White House had been attacked and President Obama was injured. By the time the veracity of this message was confirmed US stock market incurred huge losses.

Expanding Social Community: In times like these wherein social profiles of prominent users like politicians, film stars etc. are often under scrutiny, such bots aid in manipulating their social network statistics by artificially inflating their popularity.

Fake product promotion: Here the users share a fake product at global scale so as to increase its popularity. Bots have been created in online e-commerce

websites to influence the rating and garner the attention of many users towards the product.

Phishing attacks: Several fake links are shared via bots which when clicked can harvest confidential information of the users.

1.2 Types of Sybil Attacks [5]

Figure 1 categorizes Sybils in OSN.

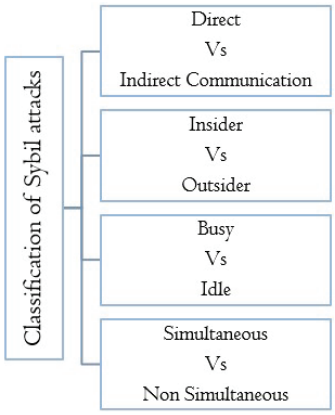

Figure 1 Classification of Sybils.

1. Direct vs. In-Direct communication
   In direct communication the attacker uses his fake identities to communicate directly with honest nodes wherein in indirect means he gleans the information using his real identity and redirects it to his aliases.
2. Busy vs. Idle
   Some Sybil nodes remain active while the others do not perform any activity post their creation.
3. Simultaneous vs. Non-Simultaneous
   In a simultaneous attack the attacker attacks with all his pseudonyms at once, whereas, if he slowly gets his fictious identities into the network, then the attack becomes non- simultaneous.
4. Insider vs. Outsider
   A person who is a part of the organization can introduce many replica nodes into the network in an easier fashion than an outsider.
   The key contributions of this research are as follows:

- An Improved Sybil Guard classifier is developed which classifies if the profile is authentic or not.
- The performance of the proposed classifier is evaluated against popular algorithms in various categories of sybil detection.

2 Related Works

Breuer et al. (2020) [6] introduced an improved graph-based detection technique namely Sybil Edge which detected the fakes by thoroughly analyzing their friend request choices. Here the authors examined how frequently such nodes sent out requests and who were the targets to such requests. The general observation noted by the authors were sybils often targeted active users and those accounts who tended to accept request from any profile, accepted such requests. However, like all graph-based techniques, the above method fails if the attacker can craftily achieve a vast network with a bunch of real users.

Gao et al. (2020) [7] incorporated an ensemble of deep learning algorithms which comprised of CNN to extract low level features and bidirectional LSTM to extricate correlated features. These features were fed into Softmax classifier to classify if the account is sybil or not. More features aided the classification process and reduced human effort. The authors aim to further improve their work by coalescing structure based neural network in their work.

Savyan PV and S. Mary Saira Bhanu (2020) [8] devised UbCadet which analyzed the discrepancies in users tweeting patterns so as to assess if the account is compromised or not. The key focus of this work is to scrutinize individual user behaviors such as changes in tweet frequency, tweet topics and hashtags along with the variation in geolocation of the tweets. The proposed technique however fails in community detection of bots. The authors hope to improvise the work by adding semantic analysis of the tweets to gain better detection.

Kumari et al. (2020) [9] devised an algorithm to detect sybil attacks in communication modules of the cyber physical systems. Sybil attacks on such systems often lead to data loss, delay and loss of packets due to packet drop or wrong path. The proposed system analyzes throughput, delay and energy along with corroborating the identity of the node and providing data forthrightness. The authors foresee in future to introduce fuzzy logic or digital signature-based components to their work.

Rheem Althari et al. (2019) [10] incorporated a Sybil detection technique comprising of Label propagation and Label spreading algorithms which were

trained on a 16-feature set listed by them. Their classifiers were trained on various hyper parameters to obtain the most optimal solution. However, it was observed that their work was limited to a specific feature set whose scope needs to be scaled.

Shu et al. (2019) [11] formulated SybSub to detect fake publishers and subscribers in a crowdsourcing environment. The proposed technique which is an amalgamation of modified Paillier homomorphic cryptosystem and the ID-based signature schemes, proved effective in inspecting numeric disparities which was the key- shortcoming of earlier keyword based matching schemes.

Feng et al. (2019) [12] generated the following five-fold feature set to find attackers in Weibo social network. They are:

- the activity level of the user.
- frequency of interactions.
- no. of users who follow the person in question.
- social circle likeness using Jaccard similarity coefficient.
- preference list of users using topic detection algorithms on their blogs.

The mentioned features were fed into various machine learning classifiers whose accuracies were compared.

Yuan et al. (2019) [13] devised a method to detect the Sybils at the time of registration in online social networks. Their method ensured that the damage caused due to delayed detection of such nodes can be averted. Their detection technique captures registration attributes of nodes in We chat platform. They discovered Sybil nodes from duplication of IPs, device IDs etc. Along with these, features like time, location, nick name motif etc. are also scrutinized. Their method aided in community detection of sybils. Crafty attackers can however dodge these detection patterns. The authors plan to incorporate unsupervised learning methods further to extend this work.

Faiza Masood et al. (2019) [14] reviewed spammers and spam detection techniques by performing an organized study wherein unsolicited content is identified via analysis of fake paraphernalia, URLs, user profile contents who tweets it and spam in trending topics. Their research gave a good feature set to analyze hoax user behavior. Some of the key research areas highlighted include identification of veracity of rumors and its source.

Muhammad Al-Qurishi et al. (2018) [15] developed a three-tier architecture model which exercises deep regression model on the ingathered features and check if a node classifies as sybil or not. The collected feature set were grouped into emotion based, profile based, graph based,

topic/temporal/quality-based categories. A feed forward neural network was fed with this feature set and it gave an accuracy of 86% despite of being subjected to a tainted dataset.

Zhang et al. (2018) [16] analyzed user's friend network and activities to assist Sybil detection. But many honest users are inactive, they are often misclassified in this approach.

Sybil nodes in MANETS often drop the packets causing packet loss. P. Muthusamy et al. (2018) [17] concocted an indigenous way of employing digital certificates to authenticate the nodes in the route to destination node. Their research impresses on the presence of sybils in all domains and also establishes the fact that behavior-based approach is suitable in detecting such dishonest nodes.

Wu et al. (2017) [18] analyzed the abnormal behavior of the sybil nodes by comparing it similarity to honest nodes based on 2 parameters. Their static similarity score is the degree of closeness based on profile attributes and preferences amongst 2 friends. Along with it, the interaction rate of the nodes is also scrutinized. The filtered nodes were tracked to identify if they flaunt any pen names or not via Hidden Markov Model. When evaluated against traditional Sybil classifiers like Sybil Guard and Sybil Defense, promising results were observed.

Wang et al. (2017) [19] developed a framework which combined the merits of Random Walks (RW) and Loop Belief Propagation (LBP). The developed framework was integrated with local rules where a neighbor's label influences the prediction. Although the developed classifier was more robust to noise and scalable when compared to traditional RW and LBP methods, more research can be done to learn the strength of each edges.

Zheng et al. (2017) [20] introduced ELSIEDET, which does the following:

- Identifies doubtful users.
- Ascertains the agenda they are working on (ex. spreading hate about a person).
- Groups all the users participating in the crusade by employing Similarity Metrix.

However, detection in this case can be evaded if the Sybil nodes become dormant.

Cresci et al. (2016) [21] incorporated DNA Analysis techniques to identify spam bots by building a profile for genuine users and identifying the

subsequent variations. More DNA (tweet type, content) must be incorporated to get better results.

Shekokar et al. (2016) [22] enriched the social graph detection by incorporating it with behavioral analysis. But, sybils could evade detection if they mimicked the behavior of honest nodes.

Samuel et al. (2015) [23] employed trust relationship and improved knowledge discovery tree to find inconsistency amongst the nodes. An improved detection when compared to the earlier trust-based approaches was observed.

Manuel Egele et al. (2015) [24] developed COMPA which statistically analyzes the online behavior patterns pertaining to certain profile to check if there is a breach or not. To build the profile the following features were streamed per user; most active time period, device information, proximity of interaction, topics tweeted, common mentions, language, styles and links used. However, the reliability of the proposed method depends on how much comprehensive data can be collected from the profile to establish a good behavioral pattern which can be used for learning.

In their work authors Gaur et al. (2015) [25] focused on analysis of images which often are left undetected by antispam filters. By applying OCR text in such images were analyzed so as to check if they are spam or not. Also, Bayesian algorithm was employed to implore on known and documented patterns to aid in detection. Futher, to strengthen the detection anti stegnographic algorithms were employed to detect potentially harmful images. The authors plan onto work on other multimedia formats like videos in future.

Mansour Alsaleh et al. (2014) [26] developed a browser plug-in which detected if an account is sybil or not based on the analysis provided by various machine learning classifier. Most of their feature set is used in our work.

Zheng et al. (2015) [27] applied Support Vector Machine to categorize the users based on their behavior and message content. The approach lacked, due to more training time spent on the classifier, which is not feasible for real world detection. Deep Learning algorithms can be incorporated for automatic feature extraction.

Zang et al. (2013) [28] developed a two-class model which first separated honest and attack nodes and then calculated the interaction probability between them. Their work was based on the hypothesis that sybils connect only with limited group of users. The authors strategize in future, to expand their work in the domain of anonymous systems and validate it against real connections.

Cao et al. (2013) [29] amalgamated the graph-based detection mechanism with user feedback. Sometimes even genuine users can get negative feedback, which compromises this approach.

Xu et al. (2010) [30] categorized the authentic and con nodes into separate clusters and tried to minimize the interaction amongst them by identifying the attack edges. Owing to large amount of data structures used, memory cost of this method is much more than its predecessors. Threshold metric used to identify an attack edge proved not accurate enough.

3 Methodology

The Figure 2 gives a detailed explanation about the proposed methodology.

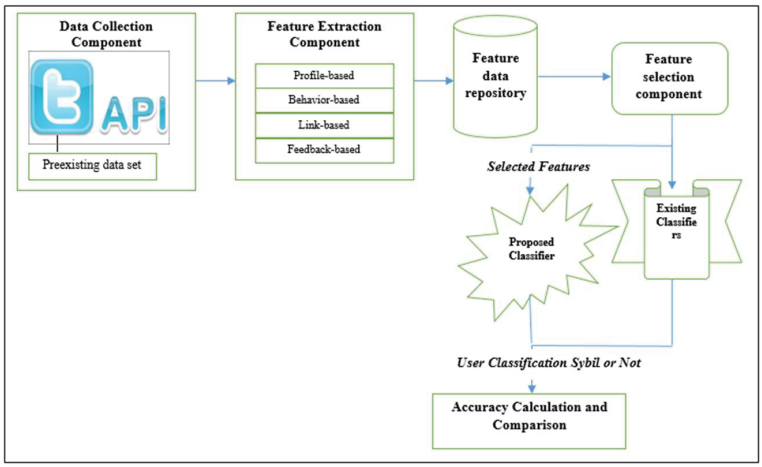

Figure 2 Overall methodology.

1. Data Collection and Feature Engineering
   Two sets of data (Pre-existing: source- Kaggle and Machine Learning-Detecting-Twitter-Bots data set [31] and streamed using twitter API) were collected to generate network centric and profile centric properties. Attributes like no. of followers, following count, tweet frequency of the user; average number of hashtags, emojis and urls mentioned which show strong correlation is included in the work. Network Centric properties convert the data in the form of a graph (.mtx) file (users are nodes and following–followers are connected using edges). Profile Centric properties were treated to missing values. It was observed that followers

and following count contribute the most to the detection of bots (bots: less followers; following count is more, non bots: more followers).

2. Feature Selection
   The selected features as shown in Table 1, can be culminated into the following four categories:
   - Profile based: – They consist of the descriptive features like screen names and bio present in user profiles.
   - Behavior-based: – They illustrate the affinity of the people from their posts and forwards. They include the syntactical features such as average number of hashtags, emojis, user-mentions, links and special characters.
   - Feedback-based: – These features highlight the opinion of users on the activities of other users in the form of likes, comments, shares etc.
   - Link based: – The links comprises of the network constructed via following and follower's circle.

Table 1 Feature set description

Features	Description
Screen name	Username of a account
Followers count	No. of accounts following a particular user
Following count	No. of accounts a user is following
Retweet count	No. of times user's tweet is retweeted
Likes count	No. of likes obtained for a tweet (per tweet)
Replies count	Count of replies obtained to a tweet (per tweet)
Emoji count	Count of emoticons per tweet
URL count	No. of links in a single tweet
Hashtag count	Count of hashtags per tweet
User Mentions count	No. of user mentions per tweet
Similarity Index	Sentiment of top 20 tweets of same topic (0 if same , 1 if dissimilar)
Text topic	Topic of top 20 tweets (0 if same, 1 if diverse)

3. Improved Sybil Guard Classifier
   The proposed model is inclusive of the following components and is illustrated in Algorithm 1.
   - Sybil Guard [32]: A network comprises of both genuine and attack nodes. Edges connecting fake nodes to honest nodes are called as attack edges. From an authentic node and suspected node m random paths are taken. The path taken from the alleged node is

said to be verified if it intersects with the path of an honest node. A node is said to be honest if its multiple paths are verified.

- Engagement Rate: It determines the outreach of a particular tweet done by the user.

$$\text{engagement rate} = (\text{totallikes} + \text{totalretweets} + \text{length(replies)})/\text{details} \quad (1)$$

Details includes average of hashtags, user-mentions, urls and emojis used along with the tweet frequency.

- User Profile Based Characterization: The extracted features where analyzed against preexisting bag of words for bots and the thresholds set to check if the profile is authentic or not.

4. The results of the developed classifier was evaluated against popular sybil detection algorithms in various categories.
5. The accuracies of the all the techniques in the determining if the profile is honest/ fake (0/1) were compared and evaluated.

Algorithm 1 Proposed algorithm

Input: Graph G (V,E): Set of Users V in the network with their attributes, Set of connections E between the users.

Output: Decision on if a node is Sybil or not.

for all nodes i ∈ V **do**

Consider H as honest nodes, S as suspect nodes and F as number of paths from h ∈ H to any particular node i ∈ V.

Calculate FQ_i in all paths f∈F

$engagement\_rate_i = ((totallikes + totalretweets + length\ (replies))_i)/details_i$

Details includes average of hashtags; user mentions;URLs and emojis used along with the tweet frequency.

if condition **then**

$S = S \cup i$

else

$H = H \cup i$

return H, S

Note: Condition must satisfy the following :

i. $FQ_i < \text{Threshold}$
ii. $(\text{twF \&\& followingC})_i > \text{Threshold}$
iii. $(\text{avg urls \&\& avg hashtags \&\& avg userm})_i > \text{Threshold}$

iv. (followersC && engagement rate)$_i$<Threshhold && (screenname$_i$ contains bag of bots)
v. Views on top 20 tweets on the same topic not similar (1) (If similar then value is 0)
vi. Range of topics tweeted on for latest 20 tweets, same (1) (If diverse then value is 0)

4 Results

The Tables 2–5 shows the obtained results.

Table 2 Comparing against structure based methods

Classifier	Accuracy (In Percent)
Sybil Guard	65
Sybil Limit[33]	68.2
Sybil Rank [34]	69
Sybil Belief [35]	78
Sybil SCAR [36]	82
Sybil Infer [37]	71.88
Sybil Defender [38]	76.3
Proposed Method	95.34

Table 3 Comparing against crowd sourcing based methods [39]

Classifier	Accuracy (In Percent)
Chinese Expert	>90
Indian Expert	85–90
Indian or Chinese Turker	<65
Chinese and Indian Turker groups	<50
Proposed Method	95.34

Table 4 Comparing against popular machine learning algorithms [40]

Classifier	Accuracy (In Percent)
Multinomial Naive Bayes	69.72
Decision Tree	83.81
Random Forest	87.85
SentiBot [40]	88.99
Bot or Not [41]	86.4
Proposed Method	95.34

Table 5 Comparing against behavior analysis algorithms [42]

Classifier	Accuracy (In Percent)
k-Nearest Neighbor + clickstream behavior analysis	77.76
Support Vector Machine + clickstream behavior analysis	80.99
Nearest Cluster + clickstream behavior analysis	88.99
Proposed Method	95.34

5 Discussion

Common misdeeds observed by such bots are identity piracy, content adulteration, dummy follower creation, inaccurate information promulgation etc. The sybil detection approaches can be broadly categorized into the following categories

1. Graph Based Schemes: Most of these schemes work under the assumption that the attack nodes cannot establish many connections with the honest nodes. However, the proposed schemes fail in detection when a new user with limited number of connections is subjected to scrutiny. Crafty attackers tend to elude such systems by generating hoax accounts which are densely connected to real users. This causes false negatives as observed in Sybil Guard scheme. Sybil Limit, another popular defense strategy cannot detect more than one malicious node at a time. Although Sybil Infer achieves low false negatives when compared to the other approaches, it fails in achieving a nominal computation overhead. Sybil Defender outperforms all the other above approaches. However, the only limitation is that in order to save time the number of random walks in reduced in this technique, which holds good only when number of attack edges are limited and the sybil community is closely associated.
2. Machine learning approaches: These methods depend on amassing a good feature set which proficiently distinguishes bots from humans. Algorithms are trained on features like no. of URLs, mentions in tweets, tweet frequency, range and sentiments of topics tweeted on, follow-following ratio etc. The accuracy of supervised machine learning algorithms bank on collection of an extensive feature set. Experiments with algorithms like SVM, failed to give effective detection when encountered with attributes on which no suitable training samples are present. However, OSNs don't give access to their databases and also resource constraints while training are the common hindrances to this

approach. Although unsupervised approaches are less complex in terms of time and resources, common algorithms in this category like k-Means are often ineffective to group such vast networks with varying dimensions, sizes and having outliers.

3. Crowdsourcing-based approaches: This was first proposed by Wang, who thought to employ human effort like Amazon Mechanical Turk to classify the profile as a bot or not. While this approach guaranteed limited false positives and negatives, it is not often favored for large OSNs as it is too costly and time-consuming to label. When this method is enforced, adequate measures must be taken to protect personal information of users, to prevent privacy leak. Proper training to the annotators is essential to ensure consistency in the work.
4. Behavior based detection: These schemes exploit the users browsing and clicking habits which constitute how frequently users expand their social circle, send messages, share contents etc. Algorithms like Markov chains have inferred that sybils tend to engage in same activity throughout and fail to diversify like real users. Such schemes fail when a crafty attacker mimics a real user and can detect only basic sybil nodes (whose feature set match with those collected in the database).
5. Sybil Prevention Methods: Common schemes such as CAPTCHA, solving some crypto puzzles, or verification via registered phone numbers/email-Ids are used to detect bots. However, although these techniques are widely used, it is easier to circumvent them but using disposable phones and training using deep learning techniques [43, 44].

6 Conclusion and Future Works

There is a wide upsurge in creation of fake accounts to influence real users and gain social support. Studies have proved that interaction (offline/online) have known to impact a person's emotion and consequently action. Vile users employ strategies like using an attractive female photo while befriending male users and subsequently utilize the person to fulfill their agendas. Sybil nodes while compromising data integrity and privacy can further launch more deadlier attacks such as phishing, DOS, etc. which must be mitigated. The proposed methodology can be further extended to detect web spams, fake reviews, fake likes etc. Sybil defense strategies can be extended to many fields like sensor networks, vehicular ad hoc networks (VANETS), Internet of things (IOTS) etc. The suggested mixture of feature-based detection and

graph-based detection technique provides a good insight on whether a node is honest or not. False negatives are very much lessened when compared to the base algorithm.

Designing novel advanced detection techniques which support huge real time data processing while managing time, effort and computational resource constraints effectively is a challenge for researchers in this field. Further research directions include prohibiting fake users from causing negative consequences and deriving reckonable outcomes while using online platforms. Establishing ways to determine the strength of each connection is useful in improving detection of fake nodes. Open problems exist in the domain of content-based analysis and detection of bots wherein the posts, likes and other such activities of the bots are examined. Measures to detect dormant/isolated bots along with active ones and the source of these sybil nodes can be sired in future. Ways to effectively enrich the data set in this domain in order to improve the detection can be actualized. Research on autonomous intelligent agent-based approaches can be a new direction in mitigating the privacy loss risk as observed in crowdsourcing based methods.

Author Contributions

The above work was conceptualized, written and reviewed by Ms. Nisha P. Shetty. Ms. Arshia Anand and Mr. Sushant Kumar implemented the methodology. Dr. Balachandra Muniyal supervised this research.

Funding

This research received no external funding.

Data Availability Statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to violation of private data of individuals when shared without their consent.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] List of most-downloaded google play applications (Dec 2020). URL – https://en.wikipedia.org/wiki/List_of_most-downloaded_Google_Play_applications.

[2] M. Al-Qurishi, M. Al-Rakhami, A. Alamri, M. Alrubaian, S. M. M. Rahman, M. S. Hossain, Sybil defense techniques in online social networks:A survey, *IEEE Access*. 5 (2017) 1200–1219. doi:10.1109/ACCESS.2017.2656635.

[3] Sybil attack (Dec 2020). URL – https://en.wikipedia.org/wiki/Sybil_attack

[4] S. Cresci, A decade of social bot detection, *Commun. ACM* 63(10) (2020) 72–83. doi:10.1145/3409116. URL – https://doi.org/10.1145/3409116

[5] R. Gunturu, Survey of sybil attacks in social networks, *CoRR* abs/1504.05522 (2015). arXiv:1504.05522. URL – http://arxiv.org/abs/1504.05522

[6] A. Breuer, R. Eilat, U. Weinsberg, Friend or faux: Graph-based early detection of fake accounts on social networks (2020). *arXiv*:2004.04834.

[7] T. Gao, J. Yang, W. Peng, L. Jiang, Y. Sun, F. Li, A content-based method for sybil detection in online social networks via deep learning, *IEEE Access* 8 (2020) 38753–38766. doi:10.1109/ACCESS.2020.2975877.

[8] S. P. Velayudhan, S. Bhanu, Ubcadet: detection of compromised accounts in twitter based on user behavioural profiling, *Multimedia Tools and Applications* (2020) 1–37.

[9] D. Kumari, K. Singh, M. Manjul, Performance evaluation of sybil attack in cyber physical system, *Procedia Computer Science* 167 (2020) 1013–1027, international Conference on Computational Intelligence and Data Science. doi: https://doi.org/10.1016/j.procs.2020.03.401. URL – http://www.sciencedirect.com/science/article/pii/S187705092030867X

[10] R. Alharthi, A. Alhothali, K. Moria, Detecting and characterizing Arab spammers campaigns in twitter, *Procedia Computer Science* 163 (2019) 248–256, 16th Learning and Technology Conference 2019. Artificial Intelligence and Machine Learning: Embedding the Intelligence. doi: https://doi.org/10.1016/j.procs.2019.12.106. URL – http://www.sciencedirect.com/science/article/pii/S1877050919321453

[11] J. Shu, X. Liu, K. Yang, Y. Zhang, X. Jia, R. H. Deng, Sybsub: Privacy-preserving expressive task subscription with sybil detection in

crowdsourcing, *IEEE Internet of Things Journal* 6(2) (2019) 3003–3013. doi:10.1109/JIOT.2018.2877780.

[12] B. Feng, Q. Li, Y. Ji, D. Guo, X. Meng, Stopping the cyberattack in the early stage: Assessing the security risks of social network users, *Security and Communication Networks* 2019 (2019) 1–14. doi:10.1155/2019/3053418.

[13] D. Yuan, Y. Miao, N. Z. Gong, Z. Yang, Q. Li, D. Song, Q. Wang, X. Liang, Detecting fake accounts in online social networks at the time of registrations, *in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19, Association for Computing Machinery*, New York, NY, USA, 2019, pp. 1423–1438. doi:10.1145/3319535.3363198. URL – https://doi.org/10.1145/3319535.3363198

[14] F. Masood, G. Ammad, A. Almogren, A. Abbas, H. A. Khattak, I. Ud Din, M. Guizani, M. Zuair, Spammer detection and fake user identification on social networks, *IEEE Access* 7 (2019) 68140–68152.

[15] M. Al-Qurishi, M. Alrubaian, S. M. M. Rahman, A. Alamri, M. M. Hassan, A prediction system of sybil attack in social network using deep regression model, *Future Generation Computer Systems* 87 (2018) 743–753. doi: https://doi.org/10.1016/j.future.2017.08.030. URL – http://www.sciencedirect.com/science/article/pii/S0167739X17300821

[16] X. Zhang, H. Xie, J. C. S. Lui, Sybil detection in social-activity networks: Modeling, algorithms and evaluations, *in: 2018 IEEE 26th International Conference on Network Protocols (ICNP)*, 2018, pp. 44–54. doi:10.1109/ICNP.2018.00015.

[17] P. Muthusamy, T. Sheela, Sybil attack detection based on authentication process using digital security certificate procedure for data transmission in MANET, *International Journal of Engineering Technology* 7(3.27) (2018) 270. doi:10.14419/ijet.v7i3.27.17891.

[18] D. Wu, S. Si, H. Wang, R. Wang, J. Yan, Social influence aware sybil detection in social networks, *in: 2017 IEEE/CIC International Conference on Communications in China (ICCC)*, 2017, pp. 1–4.

[19] B. Wang, L. Zhang, N. Z. Gong, Sybilscar: Sybil detection in online social networks via local rule-based propagation, *in: IEEE INFOCOM 2017 – IEEE Conference on Computer Communications*, 2017, pp. 1–9. doi:10.1109/INFOCOM.2017.8057066.

[20] H. Zheng, M. Xue, H. Lu, S. Hao, H. Zhu, X. Liang, K.W. Ross, Smoke screener or straight shooter: Detecting elite sybil attacks in user-review social networks, *ArXiv* abs/1709.06916 (2017).

[21] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, M. Tesconi, Dna-inspired online behavioral modeling and its application to spambot detection, *IEEE Intelligent Systems* 31(5) (2016) 58–64. doi:10.1109/MIS.2016.29.

[22] N. M. Shekokar, K. B. Kansara, Security against sybil attack in social network, *in: 2016 International Conference on Information Communication and Embedded Systems (ICICES)*, 2016, pp. 1–5. doi:10.1109/ICICES.2016.7518887.

[23] S. J. Samuel, B. Dhivya, An efficient technique to detect and prevent sybil attacks in social network applications, in: *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2015, pp. 1–3. doi:10.1109/ICECCT.2015.7226059.

[24] M. Egele, G. Stringhini, C. Kruegel, G. Vigna, Towards detecting compromised accounts on social networks, *IEEE Transactions on Dependable and Secure Computing* 14(4) (2017) 447–460. doi:10.1109/TDSC.2015.2479616.

[25] A. Gaur, R. K. Dubey, V. Ricchariya, Article: An anti-image technique for sybil detection in web data, *International Journal of Computer Applications* 121(24) (2015) 5–8, full text available.

[26] M. Alsaleh, A. Alarifi, A. M. Al-Salman, M. Alfayez, A. Almuhaysin, Tsd: Detecting sybil accounts in twitter, *in: 2014 13th International Conference on Machine Learning and Applications*, 2014, pp. 463–469. doi:10.1109/ICMLA.2014.81.

[27] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, C. Rong, Detecting spammers on social networks, *Neurocomputing* 159 (2015) 27–34. doi: https://doi.org/10.1016/j.neucom.2015.02.047. URL – http://www.sciencedirect.com/science/article/pii/S0925231215002106

[28] W. Zang, P. Zhang, X. Wang, J. Shi, L. Guo, Detecting sybil nodes in anonymous communication systems, *Procedia Computer Science* 17 (2013) 861–869, first International Conference on Information Technology and Quantitative Management. doi: https://doi.org/10.1016/j.procs.2013.05.110. URL – http://www.sciencedirect.com/science/article/pii/S1877050913002433

[29] Q. Cao, X. Yang, Sybilfence: Improving social-graph-based sybil defenses with user negative feedback, *ArXiv* abs/1304.3819 (2013).

[30] L. Xu, S. Chainan, H. Takizawa, H. Kobayashi, Resisting sybil attack by social network and network clustering, in: *2010 10th IEEE/IPSJ International Symposium on Applications and the Internet,* 2010, pp. 15–21. doi:10.1109/SAINT.2010.32.

[31] Jubins. (n.d.). Jubins/MachineLearning-Detecting-Twitter-Bots. Retrieved from https://github.com/jubins/MachineLearning-Detecting-Twitter-Bots

[32] H. Yu, M. Kaminsky, P. B. Gibbons, A. D. Flaxman, Sybilguard: Defending against sybil attacks via social networks, *IEEE/ACM Transactions on Networking* 16(3) (2008) 576–589. doi:10.1109/TNET.2008.923723.

[33] Haifeng Yu, Phillip B. Gibbons, Michael Kaminsky, and Feng Xiao. (2010). SybilLimit: a near-optimal social network defense against sybil attacks. *IEEE/ACM Trans. Netw.* 18(3), 885–898. doi:https://doi.org/10.1109/TNET.2009.2034047.

[34] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. (2012). Aiding the detection of fake accounts in large scale social online services. *In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (NSDI'12)*. USENIX Association, USA, 15.

[35] Gong, N., Frank, M., and Mittal, P. (2014). Sybil Belief: A Semi-Supervised Learning Approach for Structure-Based Sybil Detection. *IEEE Transactions on Information Forensics and Security*, 9, 976–987.

[36] B. Wang, L. Zhang and N. Z. Gong.(2017). Sybil SCAR: Sybil detection in online social networks via local rule based propagation, *IEEE INFOCOM 2017 – IEEE Conference on Computer Communications*, Atlanta, GA, 1–9, doi: 10.1109/INFOCOM.2017.8057066.

[37] George Danezis, Prateek Mittal.(2009). Sybil Infer: Detecting Sybil Nodes using Social Networks. *NDSS* 2009.

[38] Wei Wei, Fengyuan Xu, C. C. Tan and Qun Li. (2012). SybilDefender: Defend against sybil attacks in large social networks. *Proceedings IEEE INFOCOM*, Orlando, FL, 1951–1959, doi: 10.1109/INFCOM.2012.6195572.

[39] Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M.J., Zheng, H., & Zhao, B. (2013). Social Turing Tests: Crowdsourcing Sybil Detection. *ArXiv*, abs/1205.3856.

[40] J. P. Dickerson, V. Kagan and V. S. Subrahmanian. (2014). Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, Beijing, 620–627, doi: 10.1109/ASONAM.2014.6921650.

[41] Davis, C.A., Varol, O., Ferrara, E., Flammini, A., and Menczer, F. (2016). Bot Or Not: A System to Evaluate Social Bots. *Proceedings of the 25th International Conference Companion on World Wide Web*.

[42] Wang G, Tristan Konolige, Christo Wilson, Xiao Wang, Haitao Zheng, and Ben Y. Zhao. (2013). You are how you click: clickstream analysis for Sybil detection. *In Proceedings of the 22nd USENIX conference on Security*. USENIX Association, USA, 241–256.

[43] A. Thobhani, M. Gao, A. Hawbani, S. T. M. Ali, A. Abdussalam, Captcha recognition using deep learning with attached binary images, *Electronics* 9(9) (2020). doi:10.3390/electronics9091522. URL – https://www.mdpi.com/2079-9292/9/9/1522

[44] Y. Hu, L. Chen, J. Cheng, A captcha recognition technology based on deep learning, in: *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, 2018, pp. 617–620. doi:10.1109/ICIEA.2018.8397789.

Biographies

Nisha P. Shetty has acquired her bachelor and master's degree from Visvesvaraya Technological University. She is currently pursuing her doctorate at Manipal Institute of Technology, Manipal. She is working in the area of social network security.

Balachandra Muniyal's research area includes Network Security, Algorithms, and Operating systems. He has more than 30 publications in national and international conferences/journals. Currently he is working as the Professor in the Dept. of Information & Communication Technology, Manipal Institute of Technology, Manipal. He has around 25 years of teaching experience in various Institutes.

Arshia Anand has pursued her bachelor's degree in Computer and Communication Engineering branch from Manipal Institute of Technology, Manipal. Her areas of interest include Data Science and Full Stack Development. She is currently working as an Analyst in Goldman Sachs.

Sushant Kumar has pursued her bachelor's degree in Computer and Communication Engineering branch from Manipal Institute of Technology, Manipal. His areas of interests are Data Science and full-stack development. He is currently working in GE Renewable Energy as a Software Engineer.