Identity related Threats, Vulnerabilities and Risk Mitigation in Online Social Networks*

A Tutorial

Leila Bahri
Royal Institute of Technology - KTH
Stockholm, Sweden
lbahri@kth.se

ABSTRACT

This tutorial provides a thorough review of the main research directions in the field of identity management and identity related security threats in Online Social Networks (OSNs). The continuous increase in the numbers and sophistication levels of fake accounts constitutes a big threat to the privacy and to the security of honest OSN users. Uninformed OSN users could be easily fooled into accepting friendship links with fake accounts, giving them by that access to personal information they intend to exclusively share with their real friends. Moreover, these fake accounts subvert the security of the system by spreading malware, connecting with honest users for nefarious goals such as sexual harassment or child abuse, and make the social computing environment mostly untrustworthy. The tutorial introduces the main available research results available in this area, and presents our work on collaborative identity validation techniques to estimate OSN profiles trustworthiness.

KEYWORDS

Sybil accounts, OSN identity validation, Profile trustworthiness, Sybil marking

ACM Reference Format:
Leila Bahri. 2017. Identity related Threats, Vulnerabilities and Risk Mitigation in Online Social Networks. In *Proceedings of CCS ’17, Dallas, TX, USA, October 30-November 3, 2017,* 3 pages.
https://doi.org/10.1145/3133956.3136066

1 INTRODUCTION

Identity management in the realms of Online Social Networks (OSNs) is one of the most critical elements in discussing their security. The ability to reliably identify a profile on an OSN is the first building block towards ensuring the protection of both the users (and their data), and the OSN provider’s resources and reputation. Unfortunately, this ability is not easily achieved, and/or is not provided in today’s major OSNs. The looseness in obtaining a digital identity (i.e., a profile) on an OSN, where only a valid email account is required, facilitates their joining as open socializing platforms, but also makes them exposed to identity related attacks and threats. As a matter of fact, the percentage of fake accounts existing in today’s major OSNs is continuously increasing regardless of the efforts put into detecting them, and is reported to have increased from 5.5% in 2012 to about 12% in 2015 on Facebook, for instance.[1]

There is a huge body of work in the area of detecting fake accounts in OSNs, mostly under the research topic known as Sybil detection [7]. A Sybil account represents a forged fake identity that could have been one of million others created, at scale, by a bot, or manually at a smaller scale, but with generally common malicious aims, such as spreading malware, spying on users activity and/or stealing their personal information, or infecting the environment with fake content. Sybil detection algorithms aim at discriminating between fake and real accounts based on Sybils behavior and/or features in the network, or after they have demonstrated detected malicious activity.

Generally, most works on Sybil detection adopt one of two main appraoches. The first one is a graph based appraoch, which assumes that Sybil nodes exhibit different connection patterns in the social network compared to real ones. The second approach relies on a behavioral premise, assuming that Sybil nodes behave in ways that are easily differentiable from honest activity. However, Sybil accounts are getting more sophisticated and are moving towards getting social: i.e., Sybil accounts target misleading as many honest users as possible into befriending them, gaining as such their trust in the network and looking as good profiles to detection mechanisms. Moreover, when Sybils succeed at befriending honest users, the privacy of these latters is at stake as personal information that is supposed to be shared with real friends becomes available to fake entities that might have malicious intentions (such as spreading malware, engaging in sexual embarrassment or child abuse activities, etc.).

To address this issue, we have worked on exploring identity validation solutions from a social, collaborative perspective. We have explored how the collective wisdom of honest users in a social network could be explored to analyze new profiles based only on the content they contain. We have found that honest profiles exhibit content correlations that could be learned, both using crowdsourcing techniques [2][3], and using unsuprevised decentralized machine learning techniques [10] [12] [11], and that could be effectively used to flag fake profiles early on, based only on their content.

[*] All the work conducted by the author under this topic has taken place at the University of Insubria in Varese, Italy, mainly in collaboration with Prof. Elena Ferrari and to Ass. Prof. Barbara Carminati.

[1] According to statistics published on www.statisticbrain.com

CCS ’17, October 30-November 3, 2017, Dallas, TX, USA
© 2017 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-4946-8/17/10.
https://doi.org/10.1145/3133956.3136066

The objective of the tutorial is to initiate the audience to this topic of Sybil detection in OSNs, to its main techniques, and to the challenges facing it. A synthesis of the related literature will shed the light on the main adopted techniques and on the main challenges facing this research field. The main focus would be on user-centric and community-sourced identity validation techniques, as a promising approach to face social Sybil threats. The main goal would be to open new research directions and questions under this important area.

2 SYBIL DETECTION

Given the importance of the problem of detecting Sybils, the literature has a plethora of related research works. In general, these could be discussed under two main categories: 1) graph-based Sybil detection, and 2) behavior-based Sybil decetion. For the first category, the focus is on observations made from topological structures of the underlying social network graph, showing that Sybil accounts exhibit connection patterns that could be differentiable from real ones. For instance, works such as [19][18], [17], [4], [5], and [1] have leveraged on the observation that Sybil accounts tend to mix much faster in the graph compared to real accounts.

Under the second category, the focus is more on detecting behavioral features related to accounts activities in the social network for which Sybil accounts exhibit clearly differnetiable patterns compared to real users. For instance, features related to the frequency of sending friend requests and frequency of making new friends (e.g., [16]), to accounts names structures (e.g., [15]) or to the ratio of outgoing to incoming activity and clickstream in the network (e.g., [13]), or combinations of these have been explored. In addition to that, there is also a third category that adopts a hybrid approach and tries to use both graph-based and behavioral-based features, such as the works in [6], [8], or in [9].

These automated techniques for Sybil detection face challenges mostly related to the countinuous sophisitication of Sybil accounts that adapt themselves to the used features and can manage to look or to behave as normal acounts do. To overcome these limitations, the exploitation of the human factor, as a second layer of protection, remains one of the most effective solutions, such as argued in [14].

3 IDENTITY VALIDATION

Without denying the importance of automated Sybil detection techniques, collaborative identity validation is also important as another layer of protection against more sophisticated Sybil accounts. Identity validation refers to labelling accounts with metrics that represent their estimated trustworthiness in the social network, mostly as perceived by the general users. We have conducted several research works on this topic, starting by a crowd-sourcing based technique for analyzing social network profiles and detecting those looking uncoherent and suspicsious compared to normal trends exhibited in mormal profiles [2]. In [2], we have proposed a semi-supervised learning strategy to detect correlations between attributes in a profile schema. The learned correlations were later used to parse new profiles and detect the levels of coherence they exhibit compared to other real profiles among the social group they try to connect with.

In [3], we have shown how the learning of correlations between profile attributes and the evaluation of new profiles for coherence can be achieved in a privacy preserving manner, uwing anonymization techniques. Then, in [10] and [11] we have achieved the same outcome using fully unsupervised learning within a fully dencetralized architecture where every user in the social network is only aware of her/his direct friends. We have also found that profiles trustworthiness depends on the social community they belong to. That is, it might be easier for a Sybil account to sophisticate the account to generally look real, but it can be detected within given social communities, as members of each community tend to exhibit hidden local profile patterns that can rather not be easily detected by Sybils.

4 CONCLUSION

Fighting against Sybil accounts in online social networks is a dual infinite race between detector mechanisms and the designers of the attacks. Understanding and containing identity related threats and vulnerabilities, especially in open general social networks where it is easy to sign up for a new profile, remains of utmost importance in ensuring environments of reliable information and safe social computing. The aim of this tutorial is to provide an understanding of these threats, of the existing work in the field, and to highlight the possible research direction that can be undertaken.

REFERENCES

[1] Lorenzo Alvisi, Allen Clement, Alessandro Epasto, Silvio Lattanzi, and Alessandro Panconesi. 2013. Sok: The evolution of sybil defense via social networks. In *Security and Privacy (SP), 2013 IEEE Symposium on.* IEEE, 382–396.

[2] Leila Bahri, Barbara Carminati, and Elena Ferrari. 2014. Community-based Identity Validation in Online Social Networks. In *Proceedings of the 34th international conference on Distributed Computing Systems.* IEEE.

[3] Leila Bahri, Barbara Carminati, and Elena Ferrari. 2016. COIPâĂŤContinuous, Operable, Impartial, and Privacy-Aware Identity Validity Estimation for OSN Profiles. *ACM Transactions on the Web (TWEB)* 10, 4 (2016), 23.

[4] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. 2012. Aiding the detection of fake accounts in large scale social online services. In *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12).* 197–210.

[5] George Danezis and Prateek Mittal. 2009. SybilInfer: Detecting Sybil Nodes using Social Networks.. In *NDSS.* San Diego, CA.

[6] Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos, and Shiqiang Yang. 2014. Catchsync: catching synchronized behavior in large directed graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 941–950.

[7] Krishna B Kansara and Narendra M Shekokar. 2015. At a Glance of Sybil Detection in OSN. In *2015 IEEE International Symposium on Nanoelectronic and Information Systems.* IEEE, 47–52.

[8] Naeimeh Laleh, Barbara Carminati, and Elena Ferrari. 2015. Graph Based Local Risk Estimation in Large Scale Online Social Networks. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity).* IEEE, 528–535.

[9] Yixuan Li, Oscar Martinez, Xing Chen, Yi Li, and John E Hopcroft. 2016. In a World That Counts: Clustering and Detecting Fake Social Engagement at Scale. In *Proceedings of the 25th International Conference on World Wide Web.* International World Wide Web Conferences Steering Committee, 111–120.

[10] Amira Soliman, Leila Bahri, Barbara Carminati, Elena Ferrari, and Sarunas Girdzijauskas. 2015. Diva: Decentralized identity validation for social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on.* IEEE, 383–391.

[11] Amira Soliman, Leila Bahri, Sarunas Girdzijauskas, Barbara Carminati, and Elena Ferrari. 2016. CADIVa: cooperative and adaptive decentralized identity validation model for social networks. *Social Network Analysis and Mining* 6, 1 (2016), 1–22.

[12] Amira Soliman, Leila Bahri, Jacopo Squillaci, Barbara Carminati, Elena Ferrari, and Sarunas Girdzijauskas. 2016. BeatTheDIVa - Decentralized Identity Validation in OSNs. In *icde.* IEEE.

[13] Gang Wang, Tristan Konolige, Christo Wilson, Xiao Wang, Haitao Zheng, and Ben Y Zhao. 2013. You are how you click: Clickstream analysis for sybil detection. In *Proc. USENIX Security.* Citeseer, 1–15.

[14] Gang Wang, Manish Mohanlal, Christo Wilson, Xiao Wang, Miriam Metzger, Haitao Zheng, and Ben Y Zhao. 2012. Social turing tests: Crowdsourcing sybil detection. *arXiv preprint arXiv:1205.3856* (2012).

[15] Cao Xiao, David Mandell Freeman, and Theodore Hwa. 2015. Detecting Clusters of Fake Accounts in Online Social Networks. In *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*. ACM, 91–101.

[16] Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y Zhao, and Yafei Dai. 2014. Uncovering social network sybils in the wild. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 8, 1 (2014), 2.

[17] Haifeng Yu, Phillip B Gibbons, Michael Kaminsky, and Feng Xiao. 2008. Sybillimit: A near-optimal social network defense against sybil attacks. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*. IEEE, 3–17.

[18] Haifeng Yu, Michael Kaminsky, Phillip B Gibbons, and Abraham Flaxman. 2006. Sybilguard: defending against sybil attacks via social networks. In *ACM SIGCOMM Computer Communication Review*, Vol. 36. ACM, 267–278.

[19] Kuan Zhang, Xiaohui Liang, Rongxing Lu, and Xuemin Shen. 2014. Sybil Attacks and Their Defenses in the Internet of Things. *Internet of Things Journal, IEEE* 1, 5 (2014), 372–383.