Detection of Fake Profiles in Social Media
Literature Review

Aleksei Romanov, Alexander Semenov, Oleksiy Mazhelis and Jari Veijalainen
University of Jyväskylä, Finland

Keywords: Social Network Analysis, Social Media, Fake Profiles, False Identities.

Abstract: False identities play an important role in advanced persisted threats and are also involved in other malicious activities. The present article focuses on the literature review of the state-of-the-art research aimed at detecting fake profiles in social media. The approaches to detecting fake social media accounts can be classified into the approaches aimed on analysing individual accounts, and the approaches capturing the coordinated activities spanning a large group of accounts. The article sheds light on the role of fake identities in advanced persistent threats and covers the mentioned approaches of detecting fake social media accounts.

1 INTRODUCTION

Identity is an object attached to a human being, separate from him or her. A typical example is the name of a person. Another example is a passport that contains the name, birth date and place of the person, nationality, digitally captured fingerprints and a digitally stored and a photograph of the person. A third example is a private and public key adhering to a Public Key Infrastructure. In general, identity should be unique in the sense that each identifying object must only refer to at most one person. The same person might still have several identities, like a passport and a pair of keys above, or a social security number.

The real identity is verified by authorities of some nation state. A modern passport is a typical example of this. Authorities guarantee that the picture, fingerprints, name, birthdate etc. belong to the same person, i.e. certify the object attachment. At a social media site a user is usually identified by a profile. It typically contains a picture and name, possibly an address and birth date. The sites do not, however, rigorously check that the person with the identity alluded to in the profile really created and controls the profile. If this is not the case, somebody is using somebody else's identity. This is called false identity. One can also create profiles that can use freely invented names and other information that cannot be attached to any real person in any country. In this case the identity is called a faked identity. Such a profile can still contain a picture of a real person, picked e.g. randomly from the Internet.

False identities play an important role in advanced persisted threats (APT), i.e. coordinated, lasting, complex efforts at compromising targets in governmental, non-governmental, and commercial organizations. False identities are also often involved in other malicious activities, like spamming, artificially inflating the number of users in an application to promote it, etc.

A typical scenario for using false identities is using social media platforms to impersonate someone or create a fake identity to establish trust with the target, which is then exploited:

- for gathering further information for a spear phishing attack,
- mounting a spear phishing attack, or
- for directly interacting to get the information of interest.

In the sequel we consider originally authentic, but later compromised accounts as false accounts. We also call false such accounts that contain personal information, which does not belong to the person who created this account. If the account contains, invented personal details it is called a faked account

Items that are taken as identifiers must be certified by the authorities of a country of issue, recognized inside this country, and beyond its bounds with a mutual agreement with other

Romanov, A., Semenov, A., Mazhelis, O. and Veijalainen, J.
Detection of Fake Profiles in Social Media - Literature Review.
DOI: 10.5220/0006362103630369
In *Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST 2017)*, pages 363-369
ISBN: 978-989-758-246-2

countries. As every person cannot issue an identity card by its own, different institutions are responsible for issuing proper identifiers. Banks and financial institutions issue credit cards, authorities emit passports and identity cards using different standards of reliability. One of the possible ways to create unique digital identifiers for human beings is to assign a unique string of characters to a person. For example, a social security number.

Nevertheless, a person can still create an identifier for herself in the digital world. An example of this kind of identifier can be the creation of an email address or social network profile. Whereas in "cyber space" there are various identifiers that can be connected to a real person. Those are all user names (plus the relevant passwords) in different information systems, or email addresses.

Kaplan and Haenlein (2010) define social media as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content. One of the most important building block of social media sites is user identity (Kietzmann et al., 2011). Some social media sites promote usage of real identity information, however for some it is enough to be identified by a nickname. Douceur (2002) argue, that for presenting convincingly distinct identities computing environment needs logically central trusted authority which would manage identity information; which is practically impossible.

One of the most popular social media site is Facebook at the time of writing it has around 1,8 Billion users. Facebook annual report says, that 5,5% - 11,2% of worldwide monthly active users in 2013-2014 were false (duplicate, undesirable, etc.) (Facebook, 2014).

The current article focuses on the literature review of the state-of-the-art research aiming at detecting fake profiles in social media. The available approaches that we will review are either targeting on the distinguishing characteristics of individual false social media accounts along with their social connections, or on the coordinated activities involving numerous such accounts. Nevertheless, there are a number of limitations when the approaches are considered from the perspective of APT, including the assumption of large scale activities and the low negative impact of a fake account being detected, which makes them less productive when applied in the context of APT.

Authors have analysed articles on fake profiles in social media during the period 2010 – 2016 and present findings of 28 articles on this topic. The search engine that was primarily used was Google Scholar by keywords: “fake profiles”, “social media”, “social network” and “false identities”.

2 DETECTION OF FAKE PROFILES

Fake identities in social media are often used in APT cases, both to gather intelligence prior the attack, and to establish trust and deliver malware or a link to it. Such fake identities are also used in other types of malicious activities. To combat these activities, a significant body of research to date has focused on the timely and accurate detection of the presence of a fake identity in social media.

Generally, following the taxonomy in Song et al. (2015), the approaches to detecting false social media accounts can be classified into the approaches aimed analysing individual accounts (profile-based techniques as well as graph-based methods), and the approaches capturing the coordinated activities spanning a large group of accounts.

2.1 Ad-hoc or Small-scale Use of Fake Social Media Identities

A number of fake account detection approaches rely on the analysis of individual social network profiles, with the aim of identifying the characteristics or a combination thereof that help in distinguishing the legitimate and the fake accounts. Specifically, various features are extracted from the profiles and posts, and then machine learning algorithms are used in order to build a classifier capable of detecting fake accounts (Table 1).

For instance, the paper Nazir et al. (2010) describes detecting and characterizing phantom profiles in online social gaming applications. The article analyses a Facebook application, the online game “Fighters club”, known to provide incentives and gaming advantage to those users who invite their peers into the game. The authors argue that by providing such incentives the game motivates its players to create fake profiles. By introducing those fake profiles into game, the user would increase incentive value for him/herself. At first, the authors extract 13 features for each game user, and then perform classification using support vector machines (SVMs). The paper concludes that these methods do not suggest any obvious discriminants between real and fake users.

Adikari and Dutta (2014) describe identification of fake profiles in LinkedIn. The paper shows that fake profiles can be detected with 84% accuracy and 2.44% false negative, using limited profile data as input. Methods such as neural networks, SVMs, and principal component analysis are applied. Among others, features such as number of languages spoken, education, skills, recommendations, interests, and awards are used. Characteristics of profiles, known to be fake, posted on special web sites are used as a ground truth.

Chu et al. (2010) aim at differentiating Twitter accounts operated by human, bots, or cyborgs (i.e., bots and humans working in concert). As a part of the detection problem formulation, the detection of spamming accounts is realized with the help of an Orthogonal Sparse Bigram (OSB) text classifier that uses pairs of words as features. Accompanied with other detecting components assessing the regularity of tweets and some account properties such as the frequency and types of URLs and the use of APIs, the system was able to accurately distinguish the bots and the human-operated accounts.

Detecting spamming accounts in Twitter as well as in MySpace, was also the objective of the study by Lee et al. (2010). As compared with the study by Chu et al., the set of features here was expanded to cover also the number and type of connections. A number of classifiers available in Weka machine learning suite were tried, and the Decorate meta-classifier was found to provide the best classification accuracy.

In addition to, or instead of analysing the individual profiles, another stream of approaches rely on graph-based features when distinguishing the fake and legitimate accounts. For instance, Stringhini et al. (2010) describe methods for spam detection in Facebook and Twitter. The authors created 900 honeypot profiles in social networks, and performed continuous collection of incoming messages and friend requests for 12 months. User data of those who performed these requests were collected and analysed, after which about 16K spam accounts were detected. Authors further investigated the application of machine learning for further detection of spamming profiles. On top of the features used in the studies above, the authors were also using the message similarity, the presence of patterns behind the search of friends to add, and the ratio of friend requests, and then used Random Forest as a classifier.

Table 1: Profile-based methods for detecting fake social media accounts.

Reference	Ground truth	Detection method	Accuracy
Adikari 2015	Known fake LinkedIn profiles, posted on special web sites	Number of languages spoken, education, skills, recommendations, interests, awards, etc. are used as features to train neural networks, SVMs, and principal component analysis.	84% TP, 2.44% FN
Chu et al. 2010	Manually labelled 3000x2 Twitter profiles as human, bots, or cyborgs.	1. Text classification via Bayesian classifier (Orthogonal Sparse Bigram); 2. Regularity of tweets; 3. Frequency and types of URLs; the use of APIs.	100%
Lee et al. 2010	Spam accounts registered by honeypots: 1500 in MySpace and 500 in Twitter	Over 60 classifiers available in Weka are tried. Features include: i) demographics, ii) content and iii) frequency of content generation, iv) number and type of connections. The Decorate meta-classifier provided the best results.	99,21% (MySpace), 88,98% (Twitter)
Stringhini et al. 2010	Spam accounts registered by honeypots: 173 spam accounts in Facebook and 361 in Twitter	Random forest was constructed based on the following features: ratio of accepted friend requests, URL ratio, message similarity, regularity in the choice of friends, messages sent, and number of friends.	2% FP, 1% FN (Facebook); 2.5% FP, 3.0% FN (Twitter)
Yang et al. 2011a	Spam Twitter accounts defined as the accounts containing malicious URLs: 2060 spam accounts	Graph based features (local clustering coefficient, betweenness centrality, and bi-directional links ratio), neighbor-based features (e.g., average neighbors' followers), automation-based features (API ratio, API URL ratio and API Tweet similarity), and timing-based features were used to construct different classifiers.	86% TP, 0,5% FP
Yang et al. 2011b	1000 legit and 1000 fake accounts provided by Renren	Invitation frequency, rate of accepted outgoing and incoming requests, and clustering coefficient were used as features for an SVM classifier.	99%

Seeking robust features to detect spamming Twitter accounts was also the focus of the work by C. Yang et al. (2011). Graph based features and neighbor-based features were combined with automation-based features and timing-based features in order to construct four different classifiers.

A similar approach, although with a much smaller set of features were employed by Z. Yang et al. (2011) to detect fake accounts in Renren. Clustering coefficient was used as a metric reflecting the properties of the social graphs. These features were used to build a SVMs classifier that resulted in 99% correct classifications.

Papers by Cao et al. (2011) and Conti et al. (2012) likewise propose an application of graph features for the detection of fake profiles. Cao et al. (2011) base their detection on the observation that fake (Sybil) profiles typically connect to other fake profiles, rather than the legitimate ones. Thus, there is a cut between fake and non-fake subgraphs in the graph. Conti et al. (2012) base their detection method on analysis of distribution of number of friends over time. Boshmaf et al. (2016), however, claim that the hypothesis that fake accounts mostly befriend other fake accounts does not hold, and propose a new detection method, which is based on analysis features of victim accounts, i.e. those accounts, which were befriended by a fake account. Finally, Zang et al. (2013), under the assumption that the user of a Sybil account is unable to establish a large number of friendship relationships to non-Sybil nodes, proposed the use of a generative probabilistic block model to model the growth of the social network graph and identify latent groups within this graph.

Often times, the profile-based approaches overviewed above are aimed at detecting the accounts involved in spamming. Traditional spamming, however, targets a large audience of receivers, as opposed to the spearphishing campaigns common in advanced persistent threats where a single individual or a small group of recipients is targeted instead. It is therefore unclear whether these techniques, unmodified, would perform equally well when detecting fake accounts involved in an advanced persistent threat.

This limitation is partially addressed in a work by Egele et al. (2015) who, instead of characterizing the profiles of spamming accounts, attempt to detect the cases when a high-profile legitimate account is (temporarily) subverted and acts maliciously. To this end, the authors are seeking for behavioral anomalies in these accounts, by monitoring the timing and the origin of the messages, language and message topic, URLs, use of direct interaction, and geographical proximity. These are used to construct a SVM classifier based on sequential minimal optimization algorithm. The dataset was semi-manually labelled: the messages with malicious URLs within messages, abruptly changed topics, or malicious URLs within application description pages were seen as indications of compromised profiles.

The idea of detecting (dis)similarities in user behavior was also explored in the work by Egele et al. (2015). Albeit focusing on interaction over email messages rather than through social networks, the authors nevertheless strive to detect spearphishing by profiling individual email writers and then recognizing whether a new coming email does really originate from the same profile.

2.2 Coordinated and/or Large Scale Use of Fake Social Media Identities

Instead of analysing individual profiles and their connections, many researchers focus on characterizing malicious activities involving a coordinated use of numerous accounts – for instance, in the context of black markets of bots and fake accounts for online social networks. Stringhini et al. (2013) analyse Twitter follower markets. They describe the characteristics of Twitter follower markets and classify the customers of the markets. The authors argue that there are two major types of accounts who follow the "customer": fake accounts ("sybils"), and compromised accounts, owners of which do not suspect that their followees' list is increasing. Customers of follower markets may be celebrities or politicians, aiming to give the appearance of having a larger fan base, or may be cyber criminals, aiming at making their account look more genuine, so they can quickly spread malware and spam. Thomas et al. (2013) investigate black-market accounts used for distributing Twitter spam. De Cristofaro et al. (2014) analyse Facebook like farms by deploying honeypot pages. Viswanath et al. (2014) detect black-market Facebook accounts based on the analysis of anomalies in their like behavior. Farooqi et al. (2015) investigate two black-hat online marketplaces, SEOClerks and MyCheapJobs. Fayazi et al. (2015) study manipulation in online reviews.

A specific type of large-scale fake account creation campaigns is referred to as crowdturfing, the term representing a merger of two other terms, astroturfing (i.e., sponsored information dissemination campaigns obfuscated to appear spontaneous movements) and crowdsourcing. Thus,

crowdturfing is malicious crowdsourcing. Song et al. (2015) study how to detect objects of crowdturfing tasks in Twitter.

In particular, Wang et al. (2012) describe the operational structure of crowdturfing systems, by both crawling the websites used for coordinating crowdturfing campaigns, and by executing a similar, though benign campaign of their own. The authors have found these campaigns to be highly effective in hiring users, and, given the growth in their popularity, they thus pose a serious threat to security. In a subsequent paper, Wang et al. (2014) study the applicability of machine learning approaches to detect crowdturfing campaigns, and the robustness of these approaches to being evaded by the adversaries. The paper suggests that traditional machine learning can be used to detect crowdturfing workers with the accuracy of 95-99%, albeit the detection can be relatively easily evaded if the workers adjust their behavior.

Lee et al. (2014, 2015) likewise aim at developing a method for detecting crowdturfing campaigns. The classifier built by the authors was able to achieve crowdturfing task detection accuracy of 97.35%. Further, based on comparing the profiles of crowdturfing workers at Twitter against the generic Twitter user profiles, the authors constructed a classifier that detected Twitter crowdturfing users with 99.29% accuracy. The distinguishing features used by this classifier included, among others, the variability of the number of followers over time, the graph density of the worker accounts, tweeting activity, and ratio of friends and followers.

Song et al. (2015) has proposed another method for detecting crowdturfing, CrowdTarget. Rather than aiming at detecting workers, the authors focus on detecting the target objects of crowdturfing tasks (e.g., post, page, and URL). The proposed method can successfully distinguish between crowdturfing and benign tweets with the true positive rate up to 98%, even when they both come from the same account, thus making it more robust to detection evasion techniques. The following features were proven to be discriminative: (i) retweet time distribution, (ii) the ratio of the most dominant application, (iii) the number of unreachable retweeters, and (iv) the number of received clicks.

Alas, similarly to the approaches above targeting the detection of spamming campaigns, the crowdturfing detection techniques also assume the presence of a large scale activity, and are therefore hardly able to detect a small-footprint activity carried out as a part of a targeted attack.

2.3 Other Works on Fake Social Media Identities

Krombholz et al. (2015) proposes classification of social engineering attacks into physical methods (such as dumpster diving), social approaches (relying on socio-psychological techniques), reverse social engineering (attacker attempts to make victim believe that she is a trustworthy entity, and the goal is to make the victim approach attacker e.g. for help), technical approaches, and socio-technical approaches (combining approaches above).

Kontaxis et al. (2011) describe prototype of the software which aims at finding whether profile of particular user was cloned from one online social network into another by comparing characteristics of the profiles having similar characteristics among several online social networks.

Krombholz et al. (2012) propose the raising of users' awareness as the most efficient countermeasure against social media identity theft, and describes the methods for it. Authors perform focus groups research, and suggest that the users are mostly unaware of fake profiles occurrence and its consequences.

Jiang et al. (2016) surveyed more than 100 advanced techniques for detecting suspicious behaviors that have existed over the past 10 years and presented several experimentally successful detection techniques (i.e. CopyCatch, which was described in (Beutel et al., 2013)).

3 CONCLUSIONS

False identities in the form of compromised or fake email accounts, accounts in social media, fake or cracked websites, fake domain names, and malicious Tor nodes, are heavily used in APT attacks, especially in their initial phases, and in other malicious activities. Using these fake identities, the attacker(s) aim at establishing trust with the target and at crafting and mounting a spear phishing or another attack. Based on research evidence, information gathering for a spear phishing attack heavily relies on the use of social media and fake accounts therein. It is therefore important to detect, as early as possible, the presence of a fake social media account. A number of recent research works have focused on detecting such fake accounts, either by analysing the characteristics of individual profiles and their connections, or – in case of coordinated activities, by multiple fake social media accounts,

such as in the case of crowdturfing – by analysing the commonality of these activities, too.

The main shortcoming of the majority of these research works is their implicit assumption that the owners of the fake social media accounts target a large audience of followers. While such an assumption may be valid in case of traditional spamming campaigns or in case of crowdturfing, the spear phishing commonly used in APT exhibits a different pattern of targeting only a small subset of individuals, and otherwise keeping a low profile to evade detection. As a result, the proposed detection techniques often expect, e.g., a high ratio of accepted friend requests, which is unlikely in APT. This invalid assumption, as well as the availability of other evading techniques, makes it relatively easy for the attacker behind an APT to circumvent detection.

Nevertheless, some research works are aimed at detecting the use of compromised social media accounts only involving one or few accounts, making them more applicable to APT cases. By relying on anomaly detection and one-class classification, these works are able to detect when the original user of the account has been subverted (Egele et al., 2015). Unfortunately, this only works if the real account has been compromised, but fails to detect the presence of a fake account only created for information gathering and later spear phishing. It appears that rising awareness is the only effective means of detecting such fake accounts and mitigating the risks pertaining thereto. Meanwhile, future research is needed in order to elaborate methods of fake identity detection in APT that are capable of detecting individual fake accounts having low activity profile.

The contribution of this paper consists of the literature review of current research aimed at detecting fake profiles in social media from an advanced persistent threats point of view.

REFERENCES

Adikari, S., Dutta, K., 2014. Identifying Fake Profiles in Linkedin, in: PACIS 2014 Proceedings. Presented at the Pacific Asia Conference on Information Systems.

Beutel, A., Xu, W., Guruswami, V., Palow, C., Faloutsos, C., 2013. CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks, in: Proceedings of the 22Nd International Conference on World Wide Web, WWW '13. ACM, New York, NY, USA, pp. 119–130. doi:10.1145/2488388.2488400.

Boshmaf, Y., Logothetis, D., Siganos, G., Lería, J., Lorenzo, J., Ripeanu, M., Beznosov, K., Halawa, H., 2016. Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs. Comput. Secur. 61, 142–168. doi:10.1016/j.cose.2016.05.005.

Cao, Y., Li, W., Zhang, J., 2011. Real-time traffic information collecting and monitoring system based on the internet of things, in: 2011 6th International Conference on Pervasive Computing and Applications. Presented at the 2011 6th International Conference on Pervasive Computing and Applications, pp. 45–49. doi:10.1109/ICPCA.2011.6106477.

Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S., 2010. Who is Tweeting on Twitter: Human, Bot, or Cyborg?, in: Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10. ACM, New York, NY, USA, pp. 21–30. doi:10.1145/1920261.1920265.

Conti, M., Poovendran, R., Secchiero, M., 2012. 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), ASONAM '12. IEEE Computer Society, Washington, DC, USA, pp. 1071–1078. doi:10.1109/ASONAM.2012.185.

De Cristofaro, E., Friedman, A., Jourjon, G., Kaafar, M.A., Shafiq, M.Z., 2014. Paying for Likes?: Understanding Facebook Like Fraud Using Honeypots, in: Proceedings of the 2014 Conference on Internet Measurement Conference, IMC '14. ACM, New York, NY, USA, pp. 129–136. doi:10.1145/2663716.2663729.

Douceur, J.R., 2002. The Sybil Attack, in: Revised Papers from the First International Workshop on Peer-to-Peer Systems, IPTPS '01. Springer-Verlag, London, UK, UK, pp. 251–260.

Egele, M., Stringhini, G., Kruegel, C., Vigna, G., 2015. Towards Detecting Compromised Accounts on Social Networks. IEEE Trans. Dependable Secure Comput. PP, 1–1. doi:10.1109/TDSC.2015.2479616.

Facebook, inc., 2014. Facebook annual report https://www.sec.gov/Archives/edgar/data/1326801/000132680115000006/fb-12312014x10k.htm.

Farooqi, S., Ikram, M., Irfan, G., De Cristofaro, E., Friedman, A., Jourjon, G., Kaafar, M.A., Shafiq, M.Z., Zaffar, F., 2015. Characterizing Seller-Driven Black-Hat Marketplaces. ArXiv150501637 Cs.

Fayazi, A., Lee, K., Caverlee, J., Squicciarini, A., 2015. Uncovering Crowdsourced Manipulation of Online Reviews, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15. ACM, New York, NY, USA, pp. 233–242. doi:10.1145/2766462.2767742.

Jiang, M., Cui, P., Faloutsos, C., 2016. Suspicious Behavior Detection: Current Trends and Future Directions. IEEE Intell. Syst. 31, 31–39. doi:10.1109/MIS.2016.5.

Kaplan, A.M., Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. Bus. Horiz. 53, 59–68.

doi:10.1016/j.bushor.2009.09.003.

Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S., 2011. Social media? Get serious! Understanding the functional building blocks of social media. Bus. Horiz., SPECIAL ISSUE: SOCIAL MEDIA 54, 241–251. doi:10.1016/j.bushor.2011.01.005.

Kontaxis, G., Polakis, I., Ioannidis, S., Markatos, E.P., 2011. Detecting social network profile cloning, in: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). Presented at the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 295–300. doi:10.1109/PERCOMW.2011.5766886.

Krombholz, K., Hobel, H., Huber, M., Weippl, E., 2015. Advanced Social Engineering Attacks. J Inf Secur Appl 22, 113–122. doi:10.1016/j.jisa.2014.09.005.

Krombholz, K., Merkl, D., Weippl, E., 2012. Fake identities in social media: A case study on the sustainability of the Facebook business model. J. Serv. Sci. Res. 4, 175–212. doi:10.1007/s12927-012-0008-z.

Lee, K., Caverlee, J., Webb, S., 2010. Uncovering Social Spammers: Social Honeypots + Machine Learning, in: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '10. ACM, New York, NY, USA, pp. 435–442. doi:10.1145/1835449.1835522.

Lee, K., Webb, S., Ge, H., 2015. Characterizing and automatically detecting crowdturfing in Fiverr and Twitter. Soc. Netw. Anal. Min. 5, 2. doi:10.1007/s13278-014-0241-1.

Lee, K., Webb, S., Ge, H., 2014. The Dark Side of Micro-Task Marketplaces: Characterizing Fiverr and Automatically Detecting Crowdturfing. ArXiv14060574 Phys.

Nazir, A., Raza, S., Chuah, C.-N., Schipper, B., 2010. Ghostbusting Facebook: Detecting and Characterizing Phantom Profiles in Online Social Gaming Applications, in: Proceedings of the 3rd Wonference on Online Social Networks, WOSN'10. USENIX Association, Berkeley, CA, USA, pp. 1–1.

Song, J., Lee, S., Kim, J., 2015. CrowdTarget: Target-based Detection of Crowdturfing in Online Social Networks, in: Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security, CCS '15. ACM, New York, NY, USA, pp. 793–804. doi:10.1145/2810103.2813661.

Stringhini, G., Kruegel, C., Vigna, G., 2010. Detecting Spammers on Social Networks, in: Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10. ACM, New York, NY, USA, pp. 1–9. doi:10.1145/1920261.1920263.

Stringhini, G., Wang, G., Egele, M., Kruegel, C., Vigna, G., Zheng, H., Zhao, B.Y., 2013. Follow the Green: Growth and Dynamics in Twitter Follower Markets, in: Proceedings of the 2013 Conference on Internet Measurement Conference, IMC '13. ACM, New York, NY, USA, pp. 163–176. doi:10.1145/2504730.2504731.

Thomas, K., McCoy, D., Grier, C., Kolcz, A., Paxson, V., 2013. Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse, in: Proceedings of the 22Nd USENIX Conference on Security, SEC'13. USENIX Association, Berkeley, CA, USA, pp. 195–210.

Viswanath, B., Bashir, M.A., Crovella, M., Guha, S., Gummadi, K.P., Krishnamurthy, B., Mislove, A., 2014. Towards Detecting Anomalous User Behavior in Online Social Networks, in: 23rd USENIX Security Symposium (USENIX Security 14). USENIX Association, San Diego, CA, pp. 223–238.

Wang, G., Wang, T., Zhang, H., Zhao, B.Y., 2014. Man vs. Machine: Practical Adversarial Detection of Malicious Crowdsourcing Workers, in: Proceedings of the 23rd USENIX Conference on Security Symposium, SEC'14. USENIX Association, Berkeley, CA, USA, pp. 239–254.

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., Zhao, B.Y., 2012. Serf and Turf: Crowdturfing for Fun and Profit, in: Proceedings of the 21st International Conference on World Wide Web, WWW '12. ACM, New York, NY, USA, pp. 679–688. doi:10.1145/2187836.2187928.

Yang, C., Harkreader, R.C., Gu, G., 2011. Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers, in: Proceedings of the 14th International Conference on Recent Advances in Intrusion Detection, RAID'11. Springer-Verlag, Berlin, Heidelberg, pp. 318–337. doi:10.1007/978-3-642-23644-0_17.

Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B.Y., Dai, Y., 2011. Uncovering Social Network Sybils in the Wild, in: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11. ACM, New York, NY, USA, pp. 259–268. doi:10.1145/2068816.2068841.

Zang, W., Zhang, P., Wang, X., Shi, J., Guo, L., 2013. Detecting Sybil Nodes in Anonymous Communication Systems. Procedia Comput. Sci., First International Conference on Information Technology and Quantitative Management 17, 861–869. doi:10.1016/j.procs.2013.05.110.