From Reaction to Proaction: Unexplored Ways to the Detection of Evolving Spambots

Stefano Cresci,
Marinella Petrocchi
IIT-CNR, Italy
[name.surname]@iit.cnr.it

Angelo Spognardi
Dept. of Computer Science,
Università di Roma La Sapienza, Italy
spognardi@di.uniroma1.it

Stefano Tognazzi
IMT Scuola Alti Studi Lucca, Italy
stefano.tognazzi@imtlucca.it

ABSTRACT

We envisage a revolutionary change in the approach to spambot detection: instead of taking countermeasures only after having collected evidence of new spambot mischiefs, in a near future techniques will be able to anticipate the ever-evolving spammers.

CCS CONCEPTS

• **Networks → Online social networks**; • **Computing methodologies → Genetic algorithms**; • **Information systems →** *Spam detection*; *Social tagging*;

KEYWORDS

Online social networks security; proactive spam and bot detection; genetic algorithms; digital DNA; Twitter

ACM Reference Format:
Stefano Cresci, Marinella Petrocchi, Angelo Spognardi, and Stefano Tognazzi. 2018. From Reaction to Proaction: Unexplored Ways to the Detection of Evolving Spambots. In *Proceedings of The 2018 Web Conference Companion (WWW '18 Companion).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3184558.3191595

1 THE NEVER-ENDING CLASH

As part of today's ongoing socio-technical convergence, Online Social Networks (OSNs) have a profound impact on our everyday life. We increasingly rely on OSNs content in order to form our opinions, to plan activities, and to establish social relationships. One of the most striking examples of the influence that OSNs have on our societies could be witnessed during all the latest political elections. Indeed, during the 2014 Italian mayoral elections, the 2016 US presidential elections, the 2016 UK Brexit referendum, and the 2017 French presidential elections, social media played a dominant role in the electoral campaigns, often contributing to invert the foreseen electoral outcome[1]. It is not surprising that OSNs have also been exploited for maliciously influencing the public opinion [7]. One common way to achieve this goal is to employ large groups of automated (bot) accounts (henceforth *spambots*) that repeatedly spam polarized content. Worryingly, this malicious practice is pervasive: it has been witnessed in online discussions on important societal topics (e.g., politics, terrorism, immigration) [9], as well as in debates about seemingly less relevant topics, such as products on sale on e-commerce platforms and mobile applications [3]. A fascinating peculiarity of spambots is that they *evolve* over time, adopting sophisticated techniques to evade well-established detection systems [11]. As spambots became clever in escaping detection, scholars and OSNs administrators tried to keep pace (i.e., *reacted*) with more complex detection techniques. Most notably, spambot evolution still goes on: recent investigations highlight that new waves of *social spambots* are rising [1, 3, 6]. Social spambots are now mimicking legitimate behaviors and interaction patterns in OSNs better than ever before. Being almost indistinguishable from legitimate accounts, social spambots are capable of sharing (credible) fake news, inflating the popularity of OSN users, and "reshaping political debates. They can defraud businesses and ruin reputations"[2] [10]. We are still unable to deal with these issues.

Spambot detection in OSNs is thus a never-ending clash, involving the design of techniques capable of efficiently identifying ever-evolving spammers. Until now, spambot detection has always followed a *reactive* schema. As shown in Figure 1, this schema starts with an observation of suspicious behaviors in OSNs, which leads to a study of malicious activities. Such study is exploited to design new detection techniques. As soon as the new detection techniques are deployed, spambot developers tweak the characteristics of their accounts, thus evading detection. This evolution therefore requires new observations to grasp the characteristics of the evolved spambots. As a consequence of this reactive schema, scholars and OSN administrators are constantly one step behind. In turn, this means that spambots are largely left free to tamper with our online environments.

2 PROACTIVE SPAMBOT DETECTION

Times are ripe for a revolutionary change in the approach to spambot detection. Instead of taking countermeasures only after having collected evidence of new spambot mischiefs (thus following the *reactive* schema in Figure 1), we envisage a near future where techniques will be *proactive* (Figure 2) and able to anticipate the next generations of spambots. Our ambitious vision is motivated by the combination of the following three proven facts:

(i) Recently, a behavioral modeling technique inspired by biological DNA has been proposed – and successfully applied – for monitoring and detecting spambots in OSNs [2, 4]. In the so-called *digital DNA* representation, the behavioral lifetime of an OSN account is encoded as a sequence of characters, namely a digital DNA sequence.

[1] http://www.newsweek.com/full-list-russian-twitter-bots-banned-election-meddling-probe-700703

[2] https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

WWW '18 Companion, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5640-4/18/04…$15.00
https://doi.org/10.1145/3184558.3191595

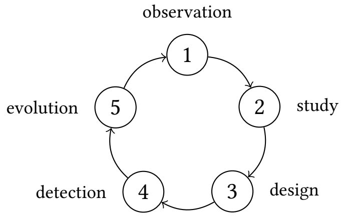

Figure 1: Reactive schema.

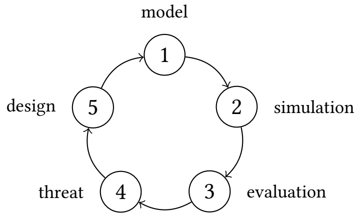

Figure 2: Proactive schema.

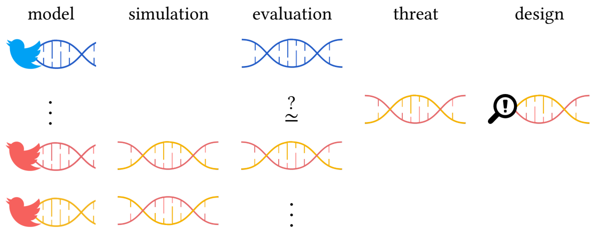

Figure 3: Proactive macro-analytical process.

(ii) Subsequently, the digital DNA sequences of legitimate and spambot accounts have been studied and compared, uncovering significant differences. In [5], authors proposed to create synthetic digital DNA sequences that resemble the characteristics of the digital DNA sequences of legitimate accounts.

(iii) Since decades, *genetic algorithms* [8] have been used as an effective heuristic to solve optimization problems. However, in order to be applied, genetic algorithms require a string-based genetic encoding of information, which severely limited their applicability when dealing with online accounts.

The combination of (i) digital DNA modeling, (ii) the possibility to create synthetic digital DNA sequences, and (iii) the evolutionary simulations allowed by genetic algorithms, open up the unprecedented opportunity to study – and even *anticipate* – the evolutionary patterns of modern spambots. In particular, the novel *proactive* schema for spambot detection is defined as in Figure 3. This process begins with modeling known spambot accounts with digital DNA. Then, genetic algorithms are exploited to create new generations of synthetic accounts with their digital DNA sequences. In this step, several *fitness functions* can be defined, in order to obtain evolved spambots with given characteristics (e.g., retweeters of other accounts, fake followers, etc.). Then, state-of-art detection techniques are evaluated against the new generations of spambots, obtained via the genetic algorithms simulations. Further, the more critical evolutions (i.e., the *threats*) of the new generations of spambots are identified, which risk to go undetected by applying the current detection techniques. Finally, novel detection techniques are designed in order to account for the newly identified threats.

Interestingly, addressing spambots evolution has already been considered before in [11], although still from a reactive viewpoint. Here, instead, the whole proactive schema can take place *before* spambots actually evolve. In other words, it does not require an *a posteriori* observation of spambot mischiefs, but instead it aims at anticipating and avoiding them.

3 EXPECTED IMPACT

For the first time since the advent of OSNs, we have the chance to proactively tackle the challenging task of spambot detection.

Although unlikely to completely defeat spambots and other malicious accounts, the application of the proposed proactive schema would nonetheless bring groundbreaking benefits. Indeed, the possibility to foresee possible evolutions of spambots, and to *a priori* design and test detection techniques, would substantially raise the bar for spambot developers. As a consequence of the additional design and experimentation allowed by the proactive schema, many spambot evolutions will be detected from day 0. Overall, spambots will have their chances to harm severely restricted, with clear and immediate benefits for our online environments, and ultimately, for our societies (e.g., less fake news). Notably, for those few spambot evolutions still not foreseen by the proactive schema, we will still be able to fall back to the traditional reactive schema. Our novel proactive schema has been here grounded on genetic algorithms and on the recent advances in digital DNA behavioral modeling, since they currently represent its key enabling factors. However, it is likely that in the future the same proactive approach to spambot detection could leverage different techniques and methodologies, thus widening the applicability of different proactive solutions.

ACKNOWLEDGEMENTS

Partly supported by the EU H2020 Program: INFRAIA-1-2014-2015: Research Infrastructures, #654024 *SoBigData: Social Mining & Big Data Ecosystem* and MSCA-ITN-2015-ETN, #675320 *European Network of Excellence in Cybersecurity (NECS)*; by Fondazione Cassa di Risparmio di Lucca (regional project *Reviewland*); by MIUR, under grant "Dipartimenti di eccellenza 2018-2022" of the Department of Computer Science of Sapienza University.

REFERENCES

[1] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2015. Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems* 80 (2015), 56–71.

[2] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2016. DNA-inspired online behavioral modeling and its application to spambot detection. *IEEE Intelligent Systems* 31, 5 (2016), 58–64.

[3] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2017. The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race. In *WWW'17 Companion*. ACM, 963–972.

[4] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2017. Social fingerprinting: detection of spambot groups through DNA-inspired behavioral modeling. *IEEE Transactions on Dependable and Secure Computing* (2017).

[5] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2017. Exploiting digital DNA for the analysis of similarities in Twitter behaviours. In *DSAA'17*. IEEE, 686–695.

[6] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[7] Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini. 2016. Detection of Promoted Social Media Campaigns. In *ICWSM*. AAAI, 563–566.

[8] Melanie Mitchell. 1998. *An Introduction to Genetic Algorithms*. MIT Press.

[9] Jacob Ratkiewicz, Michael Conover, Mark R Meiss, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2011. Detecting and tracking political abuse in social media. In *ICWSM*. AAAI, 297–304.

[10] L Steward, Ahmer Arif, and Kate Starbird. 2018. Examining Trolls and Polarization with a Retweet Network. In *WSDM'18 Workshops*. ACM.

[11] Chao Yang, Robert Harkreader, and Guofei Gu. 2013. Empirical evaluation and new design for fighting evolving Twitter spammers. *IEEE Transactions on Information Forensics and Security* 8, 8 (2013), 1280–1293.