value and opens new doors to understand the phone-based spammer ecosystem across OSNs better.

8 ACKNOWLEDGMENT

Mustaque Ahamad's participation in this research was supported in part by US National Science Foundation (NSF) grant no. CNS-1514035. We would like to thank members of Precog, IIIT-Delhi for their valuable feedback; special thanks to Paridhi Jain.

REFERENCES

[1] Hélio Almeida, Dorgival Guedes, Wagner Meira, and Mohammed J Zaki. 2011. Is there a best quality metric for graph clusters?. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 44–59.

[2] Amit A Amleshwaram, Narasimha Reddy, Sandeep Yadav, Guofei Gu, and Chao Yang. 2013. Cats: Characterizing automation of twitter spammers. In *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on*. IEEE, 1–10.

[3] Marco Balduzzi, Payas Gupta, Lion Gu, Debin Gao, and Mustaque Ahamad. 2016. MobiPot: Understanding Mobile Telephony Threats with Honeycards. In *Proceedings of the 11th ACM SIGSAC Symposium on Information, Computer and Communications Security (ASIA CCS '16)*. ACM, New York, NY, USA.

[4] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida. 2010. Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, Vol. 6. 12.

[5] Fabrício Benevenuto, Tiago Rodrigues, Virgílio Almeida, Jussara Almeida, and Marcos Gonçalves. 2009. Detecting spammers and content promoters in online video social networks. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 620–627.

[6] Juan Miguel Carrascosa, Roberto González, Rubén Cuevas, and Arturo Azcorra. 2013. Are trending topics useful for marketing. *Proc. COSN* (2013).

[7] Nicolas Christin, Sally S Yanagihara, and Keisuke Kamataki. 2010. Dissecting one click frauds. In *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 15–26.

[8] Zi Chu, Indra Widjaja, and Haining Wang. 2012. Detecting social spam campaigns on twitter. In *International Conference on Applied Cryptography and Network Security*. Springer, 455–472.

[9] Andrei Costin, Jelena Isacenkova, Marco Balduzzi, Aurélien Francillon, and Davide Balzarotti. 2013. The role of phone numbers in understanding cyber-crime schemes. In *Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on*. IEEE, 213–220.

[10] Michalis Faloutsos. 2013. Detecting malware with graph-based methods: traffic classification, botnets, and facebook scams. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 495–496.

[11] Hongyu Gao, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, and Ben Y Zhao. 2010. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, 35–47.

[12] Saptarshi Ghosh, Bimal Viswanath, Farshad Kooti, Naveen Kumar Sharma, Gautam Korlam, Fabricio Benevenuto, Niloy Ganguly, and Krishna Phani Gummadi. 2012. Understanding and combating link farming in the twitter social network. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 61–70.

[13] Chris Grier, Kurt Thomas, Vern Paxson, and Michael Zhang. 2010. @ spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*. ACM, 27–37.

[14] Payas Gupta, Mustaque Ahamad, Jonathan Curtis, Vijay Balasubramaniyan, and Alex Bobotek. 2014. *M3AAWG Telephony Honeypots: Benefits and Deployment Options*. Technical Report.

[15] Payas Gupta, Roberto Perdisci, and Mustaque Ahamad. 2018. Towards Measuring the Role of Phone Numbers in Twitter-Advertised Spam. In *Proceedings of the 13th ACM on Asia Conference on Computer and Communications Security (ASIA CCS '18)*. ACM, New York, NY, USA, 12. https://doi.org/10.1145/3196494.3196516

[16] Payas Gupta, Bharath Srinivasan, Vijay Balasubramaniyan, and Mustaque Ahamad. 2015. Phoneypot: Data-driven Understanding of Telephony Threats.. In *NDSS*.

[17] Srishti Gupta, Payas Gupta, Mustaque Ahamad, and Ponnurangam Kumaraguru. 2016. Exploiting Phone Numbers and Cross-Application Features in Targeted Mobile Attacks. In *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices*. ACM, 73–82.

[18] Jelena Isacenkova, Olivier Thonnard, Andrei Costin, Aurélien Francillon, and David Balzarotti. 2014. Inside the scam jungle: A closer look at 419 scam email operations. *EURASIP Journal on Information Security* 2014, 1 (2014), 4.

[19] Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Laura Mather. 2009. Anti-phishing landing page: Turning a 404 into a teachable moment for end users. *Conference on Email and Anti-Spam* (2009). http://precog.iiitd.edu.in/Publications_files/APWGLandingPage-Turning404intoEducation.pdf

[20] Kyumin Lee, James Caverlee, and Steve Webb. 2010. Uncovering social spammers: social honeypots+ machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 435–442.

[21] Kyumin Lee, Brian David Eoff, and James Caverlee. 2011. Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter.. In *ICWSM*.

[22] Cristian Lumezanu and Nick Feamster. 2012. Observing common spam in Twitter and email. In *Proceedings of the 2012 ACM conference on Internet measurement conference*. ACM, 461–466.

[23] Eva García Martín, Niklas Lavesson, and Mina Doroud. 2016. Hashtags and followers. *Social Network Analysis and Mining* 6, 1 (2016), 1–15.

[24] Aude Marzuoli, Hassan A Kingravi, David Dewey, and Robert Pienta. 2016. Uncovering the Landscape of Fraud and Spam in the Telephony Channel. In *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*. IEEE, 853–858.

[25] Najmeh Miramirkhani, Oleksii Starov, and Nick Nikiforakis. 2017. Dial One for Scam: A Large-Scale Analysis of Technical Support Scams. In *Proceedings of the 24th Network and Distributed System Security Symposium (NDSS)*.

[26] Federal Bureau of Investigation. 2016. TECH SUPPORT SCAM - Federal Bureau of Investigation. https://www.ic3.gov/media/2016/160602.aspx. (June 2016).

[27] Miles Osborne and Mark Dredze. 2014. Facebook, Twitter and Google Plus for breaking news: Is there a winner?. In *ICWSM*.

[28] Raphael Ottoni, Diego B Las Casas, Joao Paulo Pesce, Wagner Meira Jr, Christo Wilson, Alan Mislove, and Virgílio AF Almeida. 2014. Of Pins and Tweets: Investigating How Users Behave Across Image-and Text-Based Social Networks.. In *ICWSM*.

[29] Md Sazzadur Rahman, Ting-Kai Huang, Harsha V Madhyastha, and Michalis Faloutsos. 2012. Frappe: detecting malicious facebook applications. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*. ACM, 313–324.

[30] Bharat Srinivasan, Payas Gupta, Manos Antonakakis, and Mustaque Ahamad. 2016. Understanding Cross-Channel Abuse with SMS-Spam Support Infrastructure Attribution. In *European Symposium on Research in Computer Security*. Springer, 3–26.

[31] Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna. 2010. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference*. ACM, 1–9.

[32] Kurt Thomas, Chris Grier, Justin Ma, Vern Paxson, and Dawn Song. 2011. Design and evaluation of a real-time url spam filtering service. In *2011 IEEE Symposium on Security and Privacy*. IEEE, 447–462.

[33] Kurt Thomas, Chris Grier, Dawn Song, and Vern Paxson. 2011. Suspended accounts in retrospect: an analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 243–258.

[34] Alex Hai Wang. 2010. Don't follow me: Spam detection in twitter. In *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*. IEEE, 1–10.

[35] Steve Webb, James Caverlee, and Calton Pu. 2008. Social Honeypots: Making Friends With A Spammer Near You.. In *CEAS*.

[36] Sarita Yardi, Daniel Romero, Grant Schoenebeck, et al. 2009. Detecting spam in a twitter network. *First Monday* 15, 1 (2009).

9 APPENDIX

9.1 Regular Expressions for Data Collection

We used a curated list of 400 keywords like call, SMS, WhatsApp, ring, contact, dial, reach etc to filter relevant tweets from Twitter's Streaming API. While extracting phone numbers from the tweets, we encountered variations in representation of phone numbers, for instance the number 1-888-551-2881 can be represented as 1(888)551-2881, 1(888) 551-2881, 1.888.551.2881, or 1 888 551 2881 where all variations were being counted as different phone numbers. We filtered out this noise by post-processing the data, where a couple of regular expressions were used to obtain a valid phone number from the text obtained from each post are listed below:

```
1. ('(?<= )\d{6}-\d{3}(?= )|
(?<=\[)\d{6}-\d{3}(?=\])|(?<=\()\d{6}-\d{3}(?=\))')
2. (' (\d[\d ]{5,13}\d{2}) ')
3. ('\$ *\d+[\.]*\d+|\d+[\.]*\d+\$')
4. ('^\d+\s|\s\d+\s|\s\d+$')
```