To
The Editor
Journal of WebScience

Subject: Submission of manuscript for Journal of WebScience. Paper titled "Angel or Demon? Characterizing Variations Across Twitter Timeline of Technical Support Campaigners"

Dear Editor,

Please find our manuscript for possible publication in the Journal of Web Science. We have read the *Author Guidelines* for the journal and we hereby affirm that the content of this manuscript is an extension of a paper titled "Under the Shadow of Sunshine: Characterizing Spam Campaigns Abusing Phone Numbers Across Online Social Networks", which has been previously published at the 10th ACM Conference on Web Science (WebSci), 2017. The proceedings of the conference have been published by ACM. [1] The prior version of this article focused on identifying and characterizing phone based spam campaigns that abuse Online Social Networks (OSNs). We collected data from six OSNs and identified the campaigns that run in different countries and their modus operandi. In this submitted version, we broaden the scope of our work and study in detail one particular phone based campaign viz. Tech Support campaign. With the aim to identify spammers, i.e., the accounts which spread spam, we collected a new dataset that entails spam and legitimate posts from Twitter. We study the characteristic difference between how the campaigns differ, accounts operate, and finally implement a machine learning model that separates spammers from non-spammers. A detailed breakdown of the new content / sections that is specific to this manuscript is as follows:

- **Dataset.** We obtained a new dataset of legitimate Tech Support campaign by searching phone numbers used by verified Tech Support on Twitter. Comparing this dataset with our previous spam Tech Support campaign dataset revealed some interesting similarities and differences between the two sets of campaigns. To the best of our knowledge, we have not seen past literature that had done similar work on highlighting the differences between legitimate and spam campaigns.
- **Analysis.**
  - **Characterization.** We characterized both spam and legitimate Tech Support campaign dataset to compare the difference between the accounts involved in respective campaigns. Specifically, we looked into the way content is posted by accounts, how the accounts interact amongst each other, how phone numbers are being abused etc.
  - **Machine Learning Experiments.** We curated a list of content and account based features for user accounts and applied multiple supervised learning techniques on the spam and legitimate datasets. We achieved over 95% accuracy using cross validation technique; it highlights an efficient model that can identify spammers from non-spammers involved in phone campaigns which has not been achieved in past research.

Weaknesses mentioned in prior WebSci submission as mentioned by the reviewers

- Rigorous Analysis: There is a rich literature on spam detection on social networks but authors do not relate the present analysis to inform novel insights and consistency in the way phone-based scams operate versus other modes such as clickbaits.

---

[1] https://dl.acm.org/citation.cfm?id=3201065

Our response: Since the channels in which phone and other click-bait spams operate is different; the former being telephony and latter Web, we can't relate the analysis one-on-one. However, we have highlighted how the modus operandi is different in phone based, as compared to URL based spam (Section 4.4). In addition, we highlight the ineffectiveness of existing URL based intelligence in handling phone spam (Section 5.5).

- Methodology: It is also possible that a user may have mistakenly shared the message with phone-based spam information and later retracted by deleting that post – did authors again recrawl their messages to identify such false positives for the analyses?

  Our response: During our data collection for spam legitimate campaigns, we re-crawled all the tweets that were present in our spam dataset made by non-suspended users to ensure we did not have false positives. Please note that we can not collect data for accounts that are suspended by Twitter. We did not find any false positives in our dataset.

- Authors do not discuss the reasons why this kind of spam has a market, and what are the pros and cons in terms of security / privacy trade-off that the phenomenon offers as an case of degradation within online communication.

  Our response: Added in the discussion section (Section 8).

We believe that the aforementioned contributions specific to this manuscript account for a significant enhancement as compared to the prior publication in terms of content and novelty, and are worthy of a journal publication.

Thanks,

Srishti Gupta, Gurpreet Singh, Saksham Suri, Dhruv Kuchhal, Payas Gupta, Mustaque Ahamad, Manish Gupta, Ponnurangam Kumaraguru