Digital Drift and the Evolution of a Large Cybercrime Forum

DECISION: accept

AUTHORS: Jack Hughes and Alice Hutchings

Summary of Reviews

- Review 1: 1 (4)
- Review 2: 0 (4)
- Review 3: 2 (4)

Reviews

Review 1

TOTAL SCORE: 1 (weak accept)
Overall evaluation: 1 (weak accept)
Reviewer's confidence: 4 (high)

===== Brief paper summary =====
The paper provides a study on the drift and churns in a large forum (Hack Forums). They observed a shift in focus for the most active users over time from hacking to financially motivated topics. The study also confirms how the forum activity relies on a small set of active people (both in terms of post activity and thread creation).

===== Strengths =====
-The paper provides some insight into the macro dynamics of a forum
-The paper tries to address a complex problem related to churn and drift that is useful to understand and potentially disrupt online criminal forums

===== Weaknesses =====
-It is unclear how the paper handles users that post a single time. This is pretty common in forums and some results could be biased on that.
-Some issues with the plots make unclear the results.

===== Detailed comments for the author(s) =====
-The main concern is about the users that post only one time. From the text, it is not clear where they end up. I assume currently they fall in category 0. Many users in the forum post only one time so I am wondering what is the impact of this. In particular when comparing the first and last posts in Fig.11a and 12a. The plots are almost identical. My opinion is because category 0 is dominated by users that posted only once, thus the first and last posts coincide. I would ask the author to clarify if the paper is published.
-Fig.2 re-use the same colors for different classes e.g. 2007-2017 (similar for other colors with a 10-year shift). This makes incomprehensible the plot (e.g. I assume the blue posts in 2017 are not from 2007 but from the "new " blue 2017). I think the plot is interesting and useful but must be fixed if the paper is published.
On this line, it is interesting to see in Fig.2 how all users in 2007 churned before 2008. Any insight on that?

===== Minor comments for the author(s) =====
-Fig.9,10,11: Black on Blue is not immediately readable.

Review 2

TOTAL SCORE: 0 (borderline paper)
Overall evaluation: 0 (borderline paper)
Reviewer's confidence: 4 (high)

This is a solid paper. It certainly has potential, but there are improvements which should be made before it is published:

1) In terms of the motivation, I think the contribution could be made clearer. It currently reads as if: other scholars have done X, and now we are doing Y. But it is not made clear exactly what was missing from approach X and why this matters. Then the reader needs to know the importance of Y and that this is a good approach to address what was missing from X, and that it drives the field forward.

2) For literature, this submission seems to mainly cite computer science papers. It may be that some social scientists have engaged with equally relevant ideas around quantitative approaches and sampling in large forum datasets. In fact, quantitative social scientists think carefully about these kind of issues. Possible scholars to engage with include Benoit Dupont, Masarah Paquet-Clouston, Carlo Morselli, David Decary-Hetu, Marie Ouellet, David Maimon among others. Some of these scholars are also very well known for SNA in particular (e.g. Morselli).

3) On digital drift theory, the relevance is not made clear enough. It is noted but not explained in the intro and it is not always clear how this theory drives the research. Given the title, I think more emphasis needs to be placed on digital drift. Probably some of the background section could be moved into the introduction to better motivate the study. The background section could then be retitled "Digital Drift" and focused directly on the theoretical background that is core to the study. If Digital Drift is the theoretical base for this paper then it also should be returned to in a serious way in the Discussion, to determine how the results confirmed, challenged or extended aspects of this approach. On a related point, is the significant amount of churn found surprising or expected?

4) There are quite a number of research questions, and then quite a number of contributions. For the article as a whole, but particularly for the introduction, can there be a more singular focus to what gap this paper is trying to fill and any key outcomes therein? Any key focus around a research question should be directly linked to theory. Simply mentioning theory and then mentioning RQs is not enough unless the link is specifically made. There also seem to be a lot of results but not a focused key takeaway.

5) There is no meaningful justification for the choice of HackForums. Apart from data availability what does the analysis of this forum add to the literature? This is a relatively open and low-level cybercrime forum. There should be some discussion of external validity, in terms of how likely the findings will apply to other forums. Are they likely to hold in relation to more elite hacking or malware forums? Are they likely to hold in relation to more profit-driven carding forums?

6) How do the authors account for the possibility of individuals holding multiple nicknames/profiles (at one point or at different points in time)? If this is a widespread practice, it potentially skews the findings, and should at least be acknowledged as a limitation.

7) There are too many figures and they are generally overcomplicated and hard to follow. It would be best to focus down on a smaller number of really key figures, and the most essential and simplest way to represent each.

8) Small point: on p. 1 does the [?] mean the reference has been removed for blind review?

Review 3

TOTAL SCORE: 2 (accept)
Overall evaluation: 2 (accept)

Reviewer’s confidence: 4 (high)

This paper examens the ‘digital drift’ phenomenon on a large underground cybercrime forum. The authors present numerous interesting findings on churn, topic and user evolution.

Strengths

+ Well-structured and readable paper, that empirically investigates anecdotical evidence from the security industry - like a shift in topical discussions on forums.
+ This is a good data analysis paper that uses the crawled forums effectively to investigate a number of patterns theoretically grounded in earlier work. The patterns discussed are somewhat exhaustive with respect to the datasets considered.
+ Relatively speaking, this paper has some of the nicest, easiest to read graphs.

Weaknesses

+ The authors leave some of the potential explanations for their findings to the reader.
+ The analysis is restricted to one forum, yet more forums have a similar profile and were active in the same time frame. This leaves one to wonder how representative the results are.

Comments for the authors

I really appreciate the work that went in to this paper. The paper is structured well, and its contributions are clear. The conceptual basis is solid, and the results show interesting patterns. My only suggestion would be to the authors to elaborate on the potential explanations of their findings - why do we observe this churn pattern, topic and user evolution? Next, I would highly appreciate a section where the authors position the used dataset in the ecosystem of underground forum, to speak to the question of the representativity of their analysis.