Surveying the Hackers: The Challenges of Data Collection from a Secluded Community

Helen Thackray, Chris Richardson, Huseyin Dogan, Jacqui Taylor, John McAlaney

Bournemouth University, Bournemouth, UK.

hthackray@bournemouth.ac.uk, crichardson@bournemouth.ac.uk, hdogan@bournemouth.ac.uk, jtaylor@bournemouth.ac.uk, jmcalaney@bournemouth.ac.uk

Abstract: There are various challenges with online data collection, from participant recruitment to ensuring the integrity and representativeness of the results; and when the data is being collected from hacking communities who value privacy the challenges become far more interesting. This study was carried out as part of an ongoing PhD project into online communities involved with hacking. This paper will discuss the resources used in designing and implementing internet-based data collection, especially with hard-to-engage participants. Data discussed has been collected via quantitative online surveys. Risks include becoming the target of a cyber-attack. Previous studies have found that research of this type is not universally welcomed in such private communities with reactions ranging from wary to hostile. The findings of this paper offer examples and areas of improvement for online research methodologies as well as reinforcing the importance of social psychological research and human factors within cyber security. The results are beneficial to those wanting to conduct their own online research in challenging areas, as well as those interested in online behaviour and hacking related topics.

Keywords: Data collection, hacking, cybersecurity, social psychology, social identity

1. Background

This research is part of an ongoing PhD project which explores the individual perspective of group identities and group processes within online hacking communities. There are many types of hacker including hacktivists, ethical hackers, and crackers. This paper contends that the conflation between hackers and cybercriminals is a damaging and misleading inaccuracy. Perpetuated by media and various government or legal agencies, it is an error which risks alienating a capable and engaged community from assisting with the development of one of the largest global resources, the internet.

One of the initial challenges often highlighted in researching communities related to hacking is the importance of privacy and anonymity to members. Online resources and communities have meant that people from all over the world can search for and join online groups based on their self-categorisation or self-identification; they can also join anonymously and explore options which might not be available offline in their local community. It has even been suggested that anonymity may in fact foster stronger communal identity (Tanis and Postmes, 2005). It is also argued that the assumption that computers damage social ties (Turkle, 1999) is incorrect, as it has been observed that the Internet "strengthens existing social movements, stimulates the formation of new ones, and mobilizes sizable numbers of people for collective action," (Postmes and Brunsting, 2002:294). It has long been known that hackers exist within social groups that share expertise, support, and training within their communities (Jordan and Taylor, 1998). From interacting with these communities, including those engaged in cybercrime, a deeper understanding of the social psychological group processes can be gained.

2. Research Challenges

This research project is carried out through netnography, an ethnography via the internet (Kozinets, 2010). Whilst some of the challenges encountered are applicable to all online research, there are others which arise because of the community being investigated. With online data collection there are often warnings about participant recruitment, as well as ensuring the integrity and representativeness of the results (Granello &

Wheaton, 2004, Lefever et al, 2007). Particular community related risks for this research included the possibility of low participation, false responses, and back lash from the community, which could entail becoming the target of a cyber-attack. Previous conversations and threads on the websites clearly demonstrated that these private communities do not always welcome researchers, with reactions ranging from wary to hostile.

2.1. Identifying, Planning and Problem Solving

Initially the websites and forums that have relevant users and discussions were identified. This was done through simple web searches using key words, and later recommendations from other forums. A user account was registered with these websites (through an anonymous email), so that conversations and threads could be observed. These accounts were used for observation not interaction. Participant observation has demonstrated processes and structure do appear to be enduring within the groups for this study, where there are multiple benefits to building a reputation on the forums. It was observed that across these sites there is an almost universal process for new members. There is always the expectation that they read the specific rules for the website or forum. The most common advice given is to “lurk moar”, or spend more time observing (lurking) the group behaviours on the forum, and learn the social norms of the group. Those that do not follow this advice and break posting rules risk being penalised or permanently banned.

To investigate the potential problems with recruitment and participation a pilot study was carried out in early 2016. This online questionnaire was hosted on Google Forms for 3 weeks and shared across sub-reddits related to hacking. It was hoped the users might be more accepting on Reddit as it is a well-known popular website, offering a less intrusive entry into the private hacking community. The threads were still met with suspicion and hostility but the study had 47 participants, which was encouraging considering the limited circulation. The overall consensus was that, although wary, the majority of participants would welcome further academic research on the social processes, stereotypes and cultures that are associated with hackers.

It was particularly useful to search these forums for previous discussions on academic research. Some discussed reported research; others were instigated by researchers conducting studies. Both were very informative in terms of what the groups were interested and/or approved of. This was invaluable knowledge when it came to composing the call for participants for the online survey. The pilot questionnaire highlighted several flaws in the design, with some participants objecting to the lack of scope for different definitions or understandings of terms, and above all, the use of Google Forms which tracks and retains user information.

With regards to the risk of a cyber-attack or cyber-bullying it was unclear how real the threat would be for different enquiries. However, to avoid the possibility this research employed an overt approach – the identity of the researcher was not concealed. This has been successful in the past (Coleman, 2015), and whilst this still holds the risk of “verbal” abuse or cyber-attack, the aim was to minimise any challenge or entertainment value in finding the identity of the researcher.

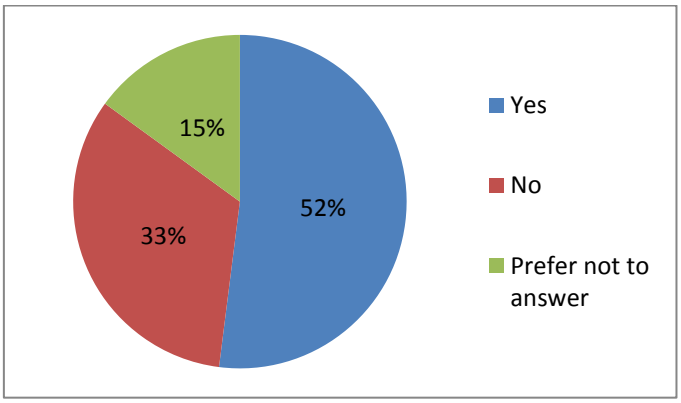
3. The Data Collection

The survey was carried out using Qualtrics, which states in its terms and conditions that all data is owned by the researcher. Included in the recruitment posts was information about the study, the site hosting the survey and the researcher. It was recommended that readers, to help keep IP addresses private, used Tor browser or a VPN connection. Again, previous approaches to these communities have led to the potential participants demanding details of the research, including ethical considerations, which were provided. To minimise the uncertainty of genuine or false information, as well as exaggeration or boasting, the questions were designed to be simple opinion on widely used terms and shared beliefs, rather than asking questions about individual experience or skill.

3.1. The Results

It was a pleasant surprise that the online survey recorded 155 submitted responses over the course of two months, shared across thirty websites and subreddits. Throughout the survey and recruitment, there were no repercussions from posting this survey, despite the inclusion of the researchers' university and topic, which made the researcher easy to find online. Feedback and responses on the forums covered the entire range of possibilities; confirmations of completion, polite and impolite refusals, and users who made clear their disapproval of the research and the presence of the researcher. Four forums banned the researcher's account entirely and deleted the recruitment post. Due to the anonymisation there is no way of knowing if anyone from these websites completed the survey.

The initial results show clearly that even those involved in the hacking communities, including forums specifically dedicated to hacking, do not necessary identify themselves as hackers. When asked "Do you consider yourself a hacker?" only 52% of participants said "yes" (see figure 1).



(Figure 1)

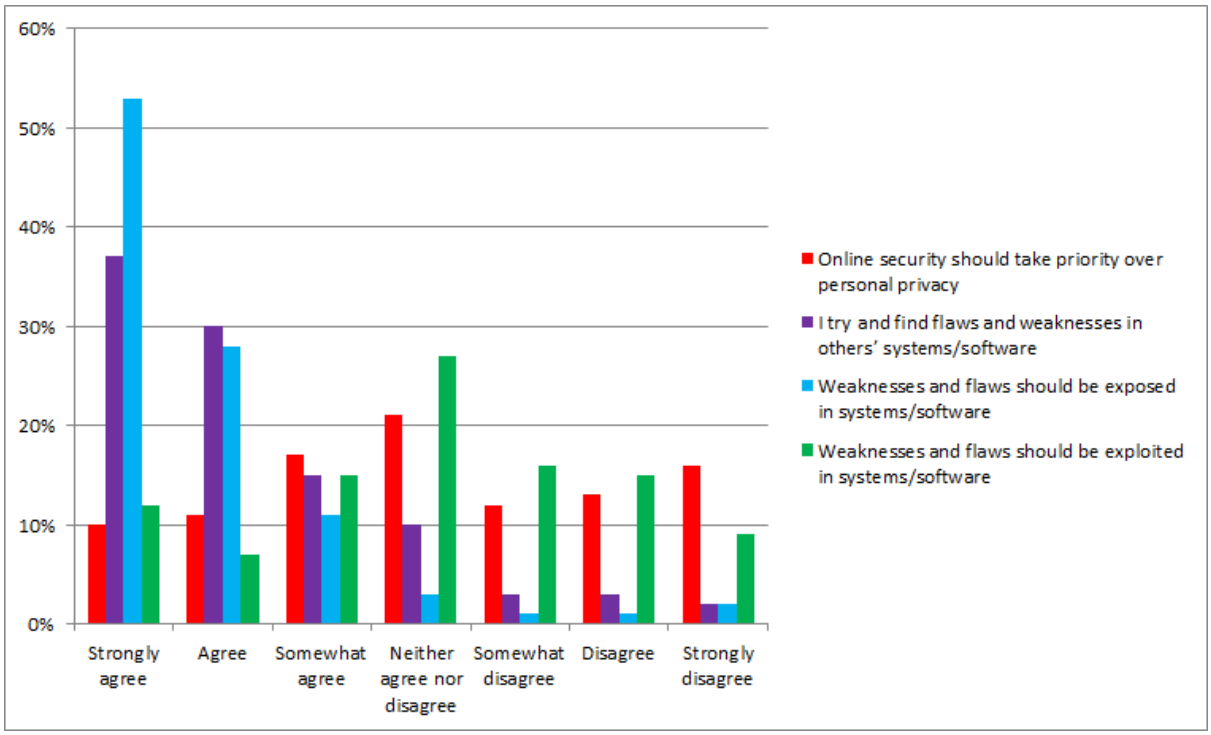
If participants answered "yes", they were then asked an additional question that was not available to other participants; they were offered different hacking sub-categories and were asked to select all that were applicable to themselves. As they could select multiple options, the results are the percentage of all participants who identified with a specific category for this question (52% of the overall participants).

Category	%
White hat hacker	38%
Black hat hacker	9%
Grey hat hacker	49%
Cracker	4%
Script kiddie	6%
Elite hacker	6%
Cyberpunk	18%
Hacktivist	13%
I disagree with these categories	11%
Other	13%

(Table 1. Selection of hacking sub-category participant belonged to)

Aside from an interesting spread of different self-categorisations, there is a mixture of assumed ethical stances, e.g. white/hacktivist = good but potentially illegal, black/cracker/script kid = bad and/or illegal. Whilst these stances are subjective, the ambiguity could explain in part why grey hat is the most chosen category. With the initial results, it is interesting that there are clear links to the "hacker ethic" (also ambiguous):

participants seem to be split relatively equally with regards to the security vs privacy debate; finding flaws and weaknesses is an active element, but what is most interesting is the sharp agreement in *exposing* problems, as opposed to *exploiting* them, where the majority were again divided across the different options, with the highest response for neutral.



(Figure 3)

In addition, at the request of many participants, the basic results of the survey were shared with these communities. This was well received, and it is believed encouraged group members to interact with the researcher further. The results have also been used as an initial discussion point in interviews.

4. Conclusions

Through using this survey as an example, it has been demonstrated that preparation and a good understanding of the target participants is essential in the success of online research. Whilst there was still abuse directed at the researcher, it was far less than expected and interspersed with other forum members defending the survey and recruitment post, citing the explanations given and the understanding of the community.

Although there are obviously many more users registered than submitted responses, given the secretive and private nature of the communities, these results are very encouraging. Although not probing deeply into the group processes, this survey has yielded some interesting considerations. Comments on the survey have reinforced that there are many users across different sites that are interested and encourage such research. This emphasises the significance and importance of social psychological research and human factors within cyber security. Through demonstrating an understanding and respect for the hacking communities' perspective, including use of appropriate terminology and acknowledging security weaknesses in the methodology, the majority of members were more willing to be participants.

References

Coleman, G. (2014). *Hacker, hoaxer, whistleblower, spy: The Many Faces of Anonymous*. London: Verso.

Davis, K., & James, C. (2013). "Tweens' conceptions of privacy online: implications for educators", *Learning, Media and Technology*, Vol. 38(1), pp. 4-25.

Granello, D. H., and Wheaton, J. E. (2004). Online data collection: Strategies for research. Journal of Counseling & Development, 82(4), 387-393.

Jordan, T. and Taylor, P. (1998) "A sociology of hackers", *The Sociological Review*, Vol. 46, pp. 757–780.

Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*. Sage publications.

Lefever, S., Dal, M. and Matthíasdóttir, Á. (2007), Online data collection in academic research: advantages and limitations. British Journal of Educational Technology, 38: 574–582.

Postmes, T. and Brunsting, S. (2002) "Collective Action in the Age of the Internet: Mass communication and online mobilization", *Social Science Computer Review, Vol.* 20-3, pp. 290-301.

Tanis, M. and Postmes, T. (2005) "Short Communication: A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour", *European Journal of Social Psychology,* Vol. 35, pp. 413-424.

Turkle, S. (1999) "Cyberspace and Identity", *Contemporary Sociology*, Vol. 28-6, pp. 643–648.