

Ethics and Internet Measurements

Jeroen van der Ham
Technical University Delft
& National Cyber Security Centre The Netherlands
Email: jeroen.vanderham@ncsc.nl

Abstract—Over the last decade the Internet has changed from a helpful tool to an important part of our daily lives for many of the world population. Where in the past it was mostly used for looking and exchanging information, it is now used to stay in touch with friends, perform financial transactions or exchange other kinds of sensitive information. This development may impact researchers performing Internet measurements as data traffic are now more likely to have some impact on users.

Traditional institutions such as Institutional Review Boards (IRBs) or Ethics Committees are not always equipped to perform a thorough review or gauge the impact of internet measurement studies. This paper examines the impact of this development for internet measurements, analyses previous cases where internet measurements have touch upon ethical issues. An early framework is also proposed to help researchers identify stakeholders and how a study may impact them.

1. Introduction

More and more daily activities involve activity over the Internet. This means that more and more data is sent over the Internet, and also that more and more *sensitive* information is sent over the Internet.

This impacts Internet measurements for cybersecurity as well, as in many cases it is hard to distinguish beforehand between sensitive and non-sensitive information. It is even possible that the results of measurements themselves become sensitive themselves.

This topic has surfaced several times in the history of internet measurement. Yet it has not received much attention so far, and deliberations on sensitivity of information, or moral aspects of experiments are often not part of the final publication.

In 2015 debates over general measurements on the Internet garnered more wider interest. There was outrage over the way Facebook performed an experiment on filtering of timeline messages [1]. And more closer to the measurement community there was debate over experiments on measurements of censorship [2], [3].

The rest of this paper is structured as follows: section 2 uses historic cases to further explore the problems surrounding measurements. Section 3 describes related work on ethical aspects of computer science. A more specific framework for measurements for cybersecurity is presented

in section 4. The paper concludes with section 5 which also contains possible future work.

2. Case Descriptions

This section describes a selection of cases that have happened over the recent years, which are in different ways relevant to the Internet Measurement community. For each of the cases the research is described first, followed by summary of the discussion and an analysis of the cause for the discussion.

2.1. Facebook filtering

In 2014 a paper was published titled ‘Experimental evidence of massive-scale emotional contagion through social networks’ [1]. This paper describes how filtering of messages with either positive or negative emotion can possibly impact the writing behaviour user of a social network. To support this conclusion an experiment had been conducted in early 2012 on more than 600.000 Facebook users.

The authors of the paper did not seek permission from the Institutional Review Board as the university researchers did not come into contact with the data itself, they only provided the analyzation methodology. Furthermore, the analysed data was a pre-existing dataset, i.e. the data had already been collected before the university researchers became involved in the research.

2.1.1. Discussion. The publication was discussed on Twitter several weeks after publication at the end of June 2014. Fairly quickly commenters discovered that Facebook had performed an experiment on unwitting users which attempted to influence their emotional state. Many participants of the discussion were outraged by how the experiment was performed and their lack of transparency on it.¹ Just several days later Cornell University issued a press release² explaining their involvement.

The discussion around this study prompted other researchers to examine the challenges around Internet-based

1. URL: <http://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/>

2. URL: <http://mediarelations.cornell.edu/2014/06/30/media-statement-on-cornell-universitys-role-in-facebook-emotional-contagion-research/>

research [4]. Their conclusion was that in 2014 in the UK there were no formal guidelines for this type of research.

2.1.2. Analysis. A study on emotions, especially when this involves a popular service such as Facebook, has a high probability of getting attention. The experiment performed on Facebook by Facebook employees is in itself legal³. However, academic research uses a higher standard for performing research. This includes reviews of human subject research by IRBs or ethics committees, and the principle of informed consent.

The IRB was not involved in this case as the data had been collected outside of the university and no university employees came into contact with the data. According to the guidelines an IRB review was not necessary for pre-existing data. This has been described by some as ‘IRB laundering’ of experimentation data [5].

The subjects in the experiment were selected by Facebook and were not approached beforehand to ask for consent. In academic practice it may be permissible to seek consent after the fact, for instance when knowledge of an experiment could influence the result. The participants in this case were not even informed afterwards about their participation, let alone about the results of the study.

2.2. Censorship Measurements

In 2015 the program committee of the ACM SIGCOMM conference had many questions on a paper describing measuring censorship [2]. So much so that the published paper includes an official statement of the program committee, calling for ethical guidelines on measurement papers.

The paper describes an experiment where certain websites are extended with measurement scripts. These measurement scripts make the browser perform certain requests for websites. This is done in such a way that it can also be recorded whether this request has been performed successfully or not.

2.2.1. Discussion. The program committee reviewing the paper found that this raised ethical concerns. The submission guidelines of SIGCOMM required authors to engage with their IRBs. The main author of the paper contacted their IRB to review the experimental approach. This IRB concluded that there were no human subjects involved in this experiment, so there was no merit for formal IRB review. The second author of the paper had moved to a different university in the meantime, so he requested a review by their IRB. Again this IRB declined a formal review because this was not human subjects research.

The paper sparked a discussion in the program committee, the measurement community and even beyond that. This case has also been extensively reviewed by Narayan and Zevenbergen [6].

3. Performing analysis of user behaviour on Facebook is in itself legal. There is discussion on the legality of using the data for scientific research since the User Agreement did not contain the ‘research’ purpose until 2014.

2.2.2. Analysis. The Institutional Review Boards of both universities seem to not be equipped to review this research. According to the statements, this is because of regulations surrounding the IRBs and the formal definition of human subject research [7], [8]. In many institutions, the IRB reviews proposed research and experiments to protect human subjects. This means an ethical review of an experiment is only done when human subjects are involved. In this case both IRBs ruled that the experiment did not touch upon human subjects.

The program committee on the other hand felt that this research did include human research subjects, or at least could possibly affect people. This shows an important disagreement between the program committee and the IRBs, which raises many different questions. There does not seem to be a way for a program committee to express concerns towards an IRB. Another question would be whether the current IRBs technically capable of accurately identifying the human subjects component of this kind of research. Another possibility is that the researchers themselves were not able to either identify the human subjects component, or were not able to explain this completely to the IRBs.

The ethical issues in this paper are again regarding informed consent, which was not performed for the experiment. The researchers argued that this was not done because it was difficult to explain, nor would it lower the risk to the user. The Program Committee was also worried about possible actions against the users browsing the websites extended with the measurement scripts. It is not unthinkable that requesting many censored websites will garner attention from law enforcement in regimes with censorship.

2.3. NYC Taxi Dataset

New York City requires that taxis keep detailed logs of their rides, including the pick-up and delivery GPS-location, as well as the price and received tip for each of the fares. This data is recorded by the city council and is thus available upon request.

A citizen researcher requested the data and received an anonymized version of the data on a USB stick. However, it turned out that the data was anonymized using simple MD5 hashing. In addition, the source data, the taxi and hack numbers did not contain enough entropy to make this a secure way of anonymization. This enabled the researcher to re-identify all of the taxi data.

To make matters worse, once the taxi numbers in the dataset were re-identified, it was also possible to correlate this to existing data. Celebrities living or visiting NYC are often photographed, also when they enter or exit taxis. Published photos often including timestamps, which makes it possible to correlate it back to a specific record in the dataset.

2.3.1. Discussion. While the case described above does not directly deal with internet traffic measurements, it does clearly show that data is not always neutral, even if it is anonymised. Anonymisation is not easy to do, especially

with low-entropy source data. Even more so when this kind of data can be correlated to other existing datasets. The taxi dataset is an obvious example of how this can go wrong.

Internet measurement data can exhibit the same qualities. For example, IPv4 addresses are very low entropy, and with a reasonably accurate timestamp this data can easily be correlated with other existing or observed datasets. Like with the taxi data, it is very hard to predict how this data can be correlated or how this would impact subjects.

2.3.2. Analysis. This case shows another area that may not currently be on the radar of IRBs. An attempt was made to anonymise the data, but apparently not enough expertise was available to do this securely. The data in the taxi set is very similar to data that is available in the measurement community: the source data has low entropy, and many related sources are available to enrich and deanonymise the original dataset.

A similar case happened in the social sciences in 2008 with the release of the ‘Taste, Ties and Times’ dataset [9]. A group of researchers published an anonymised dataset of a cohort of students at an American university. The data was very quickly deanonymised, exposing private information about students that was not generally available [10]. This study was actually reviewed by an IRB at the time, which did not see any issues in releasing this information.

2.4. Blockade Measurement

The court in The Netherlands ordered in 2012 that several ISPs had to block access to the PirateBay website. The effectiveness of this blockade was subject of debate in the court-case, so that proportionality of the measure could be established. This could not be directly measured by the ISPs themselves, as they are not allowed to inspect their own traffic.

The University of Amsterdam did perform a measurement to establish the effectiveness of the blockade. To perform the measurement it was required to collect IP addresses of downloaders, so that a distribution over the different Dutch ISPs could be established. Data was collected at moments before the blockade, during a partial blockade (two out of six major ISPs) and after an almost nation-wide blockade (all major ISPs).

2.4.1. Discussion. The ethical aspects of this research are in recording the IP addresses for the purpose of this measurement. This normally requires permission from the owner, but it also records a possibly illegal act of the downloader. At the same time, a survey was performed to measure the usage of BitTorrent by the population. Neither results are conclusive by themselves, however both gave very similar results, strengthening the conclusion that the blockade was ineffective [11]. A more extensive review of the ethical aspects of this case is described in [12].

At the time there was no IRB or ethics committee to review this experiment at the University of Amsterdam. There was no support at the university itself to guide on

this issue. The author reached out to the ethics advisor of the Twente University, who performed an ethical review of the study (after the fact).

Before this case there was already some movement to start an ethics committee for computer science at the University of Amsterdam. This case has helped shape this effort. The author has also helped start an additional ethics committee for one of the Master programs that often deals with digital security projects [13].

2.4.2. Analysis. The blockade measurement case has many different stakeholders:

- the clients downloading copyrighted materials, who are identified through this experiment;
- the copyright industry which is afraid for their income;
- the ISPs which would have to implement additional technical measures to block websites;
- the researcher performing the experiment, who must be able to demonstrate the results;
- the general public which may be harmed by increased censorship.

3. Related Work

The Menlo Report [14] was published following an initiative of the US Department of Homeland Security. The purpose of the initiative was to translate the principles of the Belmont Report to the ICT Research context, with possible additions. The companion report describes example cases and explain how ethical analyses of these cases are performed.

Partridge and Allman have made the case for an Ethical Considerations paragraph in Network Measurements publications [15]. They provide an overview of several developments in measurements publications, and call for ethical considerations paragraphs which are reviewed by program committees. The authors provide some generic advice promoting ethical awareness, but no directly applicable advice for authors or reviewers, other than referring to the previously mentioned Menlo Report.

‘Forgive us our SYNs’ [3] describes a measurement study similar to the study described in section 2.2. At the same time the paper describes an extensive overview of the ethical considerations of the study itself, aiming to provide a more generic model.

Following a Dagstuhl Seminar in 2014, several researchers cooperated to propose a model for ethics in data sharing [16]. SURFnet, the National Research and Education Network in The Netherlands, followed up with this model and created a policy for data sharing, where ethical review plays an important role [17].

4. Framework for Ethical Analysis

An earlier publication contains a description of embedding ethics in a systems and network engineering educational context [13]. The students are briefly instructed

at the start of the education on ethics. During the master programme, the students are required to write an ethical analysis for each of the projects that they perform.

For the masters program there is a (small) dedicated ethics committee that is in direct contact with the students. The ethics committee works with the students to analyse ethical aspects of their research, and also instructs them on how to improve their approach.

The model used in the education is an extension of the data sharing model as described in [16] and the Ethical Impact Assessment [18]. This means that the affected parties should be identified and the possible impact should be established. The affected parties are the users, the researcher, the university, any other related parties, as well as the general public. Many of the projects include security research, which may lead to discovery of vulnerabilities. Before the research is started, the students are forced to think about the impact of possible vulnerabilities for the vendor, the users, and the general public.

The proposed framework for internet measurements follows the same approach, with a more extensive emphasis on the different stages of measurement research:

- 1) define the purpose of the research;
- 2) design and implement the tools and experiments for the data collection and analysis;
- 3) collect the raw data (possibly by acquisition from a third party);
- 4) store the data;
- 5) analyse the data;
- 6) disseminate the results; and
- 7) curate the data.

For each of the stages the relevant parties should be identified and care should be taken for secure execution of each of the steps. It is important to emphasise that the analysis for *all of the stages* is performed before the measurement is performed.

4.1. Ethics Review

The above framework identifies the important elements of traffic measurement research. With the framework researchers can identify relevant stakeholders and start gauging an ethical impact of the research to these stakeholders. The effort of a researcher should not stop there, but they should also actively involve an IRB or EC when a direct impact is found on stakeholders other than the researcher. This should be done whether the effect is positive or negative.

The cases presented in this paper have presented evidence that computer science research, and especially traffic measurement research, can have ethical impacts. This means that IRBs and ECs should be involved to review these components of research. Therefore they should re-evaluate their regulations and procedures to gauge whether they currently can perform the reviews that are required.

Researchers themselves must also be educated on the possible impact that their research may have. The cases in

this paper show that research has changed so that measurement experiments and data can have significant ethical and legal impacts. Many of the current researchers have had only limited education on ethics, and may not realise the impact their research may have.

At the same time, institutions should take a pro-active role in supporting researchers in their ethics review. Having an ethics advisor available to discuss an experimental design is extremely helpful. In many cases it can even help to design the experiment in such a way that it has no or minimal ethical impact on other stakeholders. The Ethics Feedback Panel⁴ may also help, although it currently seems dormant.

5. Conclusion

This paper contains a description of developments and significant cases of ethical issues related to internet measurement.

- The Facebook experiment shows that informed consent is an important value in performing research.
- The censorship measurement case shows that IRBs should be better equipped to handle these cases, both by improving the regulation as well as the technical capabilities of IRBs.
- The Taxi dataset shows that sharing of data is a difficult issue. Anonymisation is hard, and impact of reidentification can not be clearly established when data is shared.
- The blockade measurement case shows that legal issues should be taken into account when performing measurements.

The existing frameworks as summarised in section 4 already can help researchers identify ethical issues in their research. Many of the previously proposed models before have been created by researchers with foresight, seeing that ethical issues exist in measurement and ICT research. These researchers worked from experience in different fields, but also from their own experience. The models indicate that ethical issues in ICT research can have a much wider impact.

The cases presented in this article show that current oversight is lacking. IRBs and ECs often do not view research necessary to review if they don't involve physical human subjects. Even if they do, these committees are often not equipped to identify human subjects in computer science research. This can either be due to lack of knowledge in these committees, or due to narrowly scoped regulations.

The Taxi dataset case shows that it is still hard to release anonymised datasets. This has been shown also for a Facebook dataset and is applicable to traffic measurement data as well. For many datasets the underlying data has too little entropy, and too much contextually relevant public data is available. Reviewing public release of datasets is also a role for IRBs and ECs, which they may not be currently equipped to handle.

4. URL: <https://ethicalresearch.org>

The cases described in this paper have created public debate, among the general public, or among the measurement community. These debates have shown that previous models identified many of the right issues, but are unfortunately still not adopted by current researchers and committees. These public debates should make clear that researchers have been ignoring for too long that ethical review of research is necessary. This ethical review should not just be applied for cases that will spark a public debate, but for other research as well. Even perfectly well-intended research can go suddenly awry and create an enormous impact on unwitting users. Academic research should strive to prevent this as much as possible.

5.1. Future Work

The related work shows that awareness among researchers is still not optimal. In recent years, the discussion has surfaced at many different measurement conferences and workshops. It is important that the community band together and create an acceptable way of identifying and acting on ethical issues in measurement research. The risk is that if this does not happen, restrictions could be imposed from the outside.

At the same time, awareness and expertise should be raised in IRBs and Ethical Committees. These IRBs and ECs should get mandate to perform reviews even when human subjects can not directly be identified. They should also proactively work with internet measurement researchers so that together they can identify possible ethical issues in the project.

References

- [1] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 24, pp. 8788–8790, 2014. [Online]. Available: <http://www.pnas.org/content/111/24/8788.abstract>
- [2] S. Burnett and N. Feamster, "Encore: Lightweight measurement of web censorship with cross-origin requests," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4, pp. 653–667, 2015.
- [3] J. R. Crandall, M. Crete-Nishihata, and J. Knockel, "Forgive us our sins: Technical and ethical considerations for measuring internet filtering," in *NS Ethics SIGCOMM*, 2015, p. 3.
- [4] S. Harriman and J. Patel, "The ethics and editorial challenges of internet-based research," *BMC Medicine*, vol. 12, no. 1, p. 124, 2014. [Online]. Available: <http://dx.doi.org/10.1186/s12916-014-0124-3>
- [5] C. Schroeder, "Why can't we be friends: A proposal for universal ethical standards in human subject research," *J. on Telecomm. & High Tech. L.*, vol. 14, p. 409, 2015.
- [6] A. Narayanan and B. Zevenbergen, "No encore for encore? ethical questions for web-based censorship measurement," *SSRN Electronic Journal*. [Online]. Available: <http://dx.doi.org/10.2139/ssrn.2665148>
- [7] U. D. of Health and H. Services, "Federal policy for the protection of human subjects (common rule)," 2014. [Online]. Available: <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>
- [8] —, "Protection of human subjects," 2009. [Online]. Available: <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html>
- [9] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, "Tastes, ties, and time: A new social network dataset using facebook.com," *Social Networks*, vol. 30, no. 4, pp. 330 – 342, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378873308000385>
- [10] M. Zimmer, "but the data is already public: on the ethics of research in facebook," *Ethics and information technology*, vol. 12, no. 4, pp. 313–325, 2010.
- [11] J. Poort, J. Leenheer, J. van der Ham, and C. Dumitru, "Baywatch: Two approaches to measure the effects of blocking access to the pirate bay," *Telecommunications Policy*, vol. 38, no. 4, pp. 383 – 392, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0308596113002152>
- [12] A. van Wynsberghe and J. van der Ham, "Ethical considerations of using information obtained from online file sharing sites," *Journal of Information, Communication and Ethics in Society*, vol. 13, no. 3/4, p. 256267, Aug 2015.
- [13] J. van der Ham, "Embedding ethics in system administration education," *The USENIX Journal of Education in System Administration*, p. 1, 2015.
- [14] D. Dittrich and E. Kenneally, "The menlo report: Ethical principles guiding information and communication technology research," U.S. Department of Homeland Security, Tech. Rep., Aug 2012.
- [15] C. Partridge and M. Allman, "Ethical considerations in network measurement papers," *Communications of the ACM*, vol. 59, no. 10, pp. 58–64, 2016.
- [16] S. Dietrich, J. Van Der Ham, A. Pras, R. van Rijswijk Deij, D. Shou, A. Sperotto, A. Van Wynsberghe, and L. D. Zuck, "Ethics in data sharing: developing a model for best practice," in *Security and Privacy Workshops (SPW), 2014 IEEE*. IEEE, 2014, pp. 5–9.
- [17] R. van Rijswijk-Deij, "Ethics in Data Sharing: a best practice for NRENs," in *Proceedings of TNC 2015*. Porto, Portugal: GÉANT, 2015. [Online]. Available: <https://tnc15.terena.org/getfile/1869>
- [18] M. Bailey, E. Kenneally, and D. Dittrich, *A Refined Ethical Impact Assessment Tool and a Case Study of Its Application*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 112–123. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-34638-5_10