Addressing Ethical Considerations in Network Measurement Papers

Craig Partridge
Raytheon BBN Technologies
craig@aland.bbn.com

Mark Allman
ICSI
mallman@icir.org

ABSTRACT

In this paper we explore the benefits of requiring measurement papers to include an ethical considerations section. We focus our attention on what specific questions such a section should answer and how to keep the section simple and easy to write for measurement studies that do not raise important ethical issues, while seeking to ensure that important ethical issues are exposed.

1. INTRODUCTION

A conference program committee is usually the first outside organization to evaluate research work in network measurement. In recent years, questions about whether the work within a particular submission has followed sound ethical practices have become more common within program committee discussions. The authors have experience as members and leaders of program committees struggling with ethics issues. The fundamental cause of this struggle is that our community does not have a set of shared ethical norms. Therefore, often the authors work from one set of ethical notions while the PC applies one or more different sets of ethical underpinnings as part of their review. Even conferences that deal with ethics in their call for papers do so in a nebulous fashion that does not make the struggle less daunting (e.g., IMC's admonishment that authors should act ethically). This leaves well meaning community members—in all roles—on fundamentally different pages. The situation is further exacerbated because our community does not have a culture of using valuable paper real estate to describe the ethical reasoning behind a set of experiments. This situation (i) leaves program committees to try to derive the foundations on which the paper stands and (ii) means that precautions taken by a careful researcher are not exposed to others who may leverage or build upon previous techniques in subsequent work.

In this paper we advocate for helping authors and program committees to get on the same page via an "ethical considerations" section in measurement papers that asks authors to answer a small number of specific questions about their work. By explicitly requiring such a section—even if a short statement that there are no ethical issues—we at least provide the starting point for a discussion about ethics in that (i) authors have a chance to justify the ethical foundations of their experimental methodologies and (ii) program committee members gain the authors perspective and can provide specific feedback as necessary. Further, by including these sections in published papers the entire community starts to develop a collective understanding of both what is ethically acceptable and how to think through ethics issues.[1]

Our aim in this short paper is to present a reasoned and initial strawman. We do not attempt to prescribe what is and what is not ethical. We do not tackle all possible ethical questions that arise in our work as Internet empiricalists. Rather, we advocate for a framework to help the community start an explicit conversation about the largest ethical issues involved in measuring the Internet.

2. BACKGROUND

There are three strands of intellectual activity that come together when one examines ethics and network measurement.

The first strand is the evolving field of ethics in information and communication.

The second strand is our ever-increasing abilities in network measurement. From simple beginnings, the field has developed a wide range of ways to extract ever more information from measurements—to the point of starting to invalidate historic divisions between kinds of measurements.

The third strand is the legal issues surrounding network measurement, which is a topic still in its infancy [7]. These issues are currently difficult in a single jurisdiction, let alone when a measurement study crosses many jurisdictions. We do not advocate engaging in illegal activities, but view these issues as independent of ethics and therefore will not discuss them further in this paper.

2.1 Ethics

The study of ethics in information and communication science has, broadly, followed two (overlapping) lines of thought and inquiry.

The first line is a focus on human-centered values such as life, health, security and happiness. This thinking originated in the 1940s with Norbert Weiner and has carried down to the present. Its best current expression is the *Menlo Report*, a 2012 report by the US Department of Homeland Security[4]. The Menlo Report focuses on issues of causing harm to persons, and, insofar as any harm may occur, that the risks of harm are recognized, moderated and equitably distributed.

The other line of ethical thinking has focused on the professional responsibility of the computing and information sciences professional. Specifically the focus has been on following good industry practices in the creation of software artifacts, and codes of conduct that outline a professional's responsibilities to society, employer, colleagues and self. The most detailed expression of this thinking that we know of is the joint IEEE/ACM *Software Engineering Code of Ethics and Professional Practice* (hereafter "the IEEE/ACM Code"), which identifies eight principles related to the Public, Client and Employer, Product, Judgment, Management,

[1] The issue of how to expose ethical issues raised by rejected papers remains. While we do not mean to diminish this important aspect of the problem, we are setting this issue aside in an effort to make some progress.

Profession, Colleagues, and Self, and presents over 80 distinct admonitions[1].

Much of the IEEE/ACM Code speaks to the day-to-day obligations of a software engineer, working in industry. Its applicability, therefore, to a researcher, is limited. But there are some clauses that have relevance, especially regarding respecting the law, disclosing possible harm, respecting the cultural milieu, respecting privacy, and ensuring that data is accurate and properly protected.

Observe that both approaches are heavily concerned with the impact of one's work on others, though the IEEE/ACM Code emphasizes an obligation to inform, while the Menlo report focuses on consent. The code places also more emphasis on privacy and confidentiality, the need to protect data, and the obligation to disclose potentially harmful results and outcomes.

We build on both approaches below.

2.2 Evolution of Network Measurement

The field of network measurement—broadly defined—is relatively old. While, as best we can tell, the early 19^{th} century optical telegraph networks were not subject to measurement[5] all subsequent networks, beginning with the electronic telegraph networks have been the subject of various forms of measurement. By 1911, AT&T had a statistical group that, among other functions, leveraged measurement to better engineer the telephone system and to predict demand.[2]

Unsurprisingly, much of our legal, social and ethical dialog about network measurement uses terminology that was developed in the early days of measurement. Specifically, the ethics and legality of network measurements largely assumes that only parties who can effectively capture data are communications companies providing service and Government agencies given access to communications companies' data centers. Further, a typical formulation distinguishes between two classes of data, as follows.

The first class of data reveals when and how long two parties communicated. United States law defines a device capable of capturing such data as a *pen register*.[3] More recently, the term *metadata* has been used to describe an expanded set of information, including packet headers, that it is argued is comparable to pen register data.

The second class of data reveals the contents of the conversation. To highlight the distinction, consider a phone call to a bank. A pen register records that a call took place at a specific time and for a specific duration. The contents of the conversation would reveal that the call was, for example, a balance enquiry. United States law, since 1967, has recognized that the contents of a conversation are a distinct class of information that has a higher expectation of privacy.[4].

We suggest that a variety of factors, including the digitization of communications, the increasing distributed control of communications, and the rise of programmable devices attached to our networks, have eroded these distinctions. Specifically, our ability to leverage metadata to infer—or even re-create—content is increasing rapidly.

A few examples illustrate this point:

- Simply measuring when devices in a network transmit (without looking at headers or any metadata) is sufficient to derive traffic tables that show which nodes are routers and which nodes are end systems and which nodes are communicating with which other nodes[3].

- The *The Queue Inference Engine* takes information about transactions (e.g. pen register style data) and reverse engineers it to determine the behavior of queues[6]. Since its invention by Larson, researchers have made steady progress in using techniques such as QIE to characterize queues from metadata. For instance, we can tell whether and roughly how long a person likely waited in line at a bank ATM machine[2].

- Interpacket gaps (metadata) between encrypted transmissions can be used to infer where users' fingers were on the keyboard and thus give guidance about what letters are in their passwords[8].

Summarizing, with less data than a pen register would collect, we can often tell who is talking to whom. If we have pen register data, we can make inferences about the environment around the participants. If we have slightly more data (metadata) we can extract passwords typed over the network, the most confidential of content.

A worrying possibility is that a number of researchers have reported informally that, during their experiments, they were able to observe timing perturbations caused by cross-traffic. These perturbations were coherent enough that the researchers could begin to make inferences about the cross-traffic. If techniques to analyze this imprinting becomes repeatable, then the analyses above will become possible *without the measuring party having access to the devices and links the measured traffic traverses*. This is particular disturbing as studies have repeatedly shown that a relatively small number of well-placed nodes is sufficient to cover most Internet links.

3. THE CONTOURS OF HARM

While there are myriad ethical issues that confront network measurement work, our aim in this paper is to address those causing *tangible harm to people*. We are not concerned with notions of potential harm to network resources (e.g. bandwidth) or equipment, except insofar as the impact on resources and equipment causes tangible harm to a human. equipment. We believe how our work impacts individual human beings is the most important ethical issue we confront and hence we tackle it first.

Additionally, we note that our goal—which agrees with the Belmont and Menlo reports—is not to eliminate the possibility of harm within our experiment. Rather, we aim to minimize both the risk of inflicting harm and the tangible harm inflicted on people. In this context we make several observations which bear on how we manage risk in our experiments:

Defining Harm: First, we recognize that "harm" is a difficult to define. Rather than a precise definition we offer that a single ICMP echo request to an IP address constitutes at best *slight harm*.[5] Meanwhile, a persistent high-rate series of probes to a given IP address may well be viewed as both an attack and create *serious harm* (e.g., by clogging a link precisely when it is needed for an emergency). These ends of the spectrum

[2] Walter S. Gifford, later President of AT&T, started his career at AT&T as its Chief Statistician in 1911.

[3] The term has an odd history. It was invented by Samuel Morse to capture *all* transmissions on a telegraph wire, but once the telephone was invented, the register was only able to capture the telephone number dialed, thus its association with connections rather than content.

[4] U.S. Supreme Court, *Katz v. United States*, 1967

[5] Of course, we have experience with complaints about these sorts of probes, which indicates that some people do in fact view them as harmful.

are useful as touchstones when thinking about how to cope with the risk involved in specific experiments.

Indirect Harm: We first recognize that the field of network measurement — for the most part — focuses on understanding *systems* and not directly assessing *people*. Therefore, any impact to people is a side effect of our measurements. While we must grapple with the ethics of harm caused by our measurements regardless of whether the harm is direct or indirect, the nature of the harm can sometimes dictate the manner in which we cope.

Potential Harm: Next we note that most often our work does not cause harm, but rather only sets up the possibility of harm. That is, additional events or factors beyond our measurements must happen or exist for actual harm to be inflicted. Again, this does not absolve us from understanding the ethics involved, but does speak to how we may manage the risk involved in conducting a particular experiment.

We believe that while fuzzy, the above aspects of "harm" offer the broad contours of the issues with which researchers must grapple. Further, we do not believe there is some one-size-fits-all way to manage harm and we allow for honest disagreement among researchers about when potential indirect harm rises to the level of making an experiment problematic. For instance, in the context of the example above about probes causing slight vs. serious harm, we discussed between ourselves whether the flooding periods could be made short enough to reasonably be felt to avoid potential harm. We agreed it was possible, but disagreed about when the experiment transitioned from slight harm to serious harm.

We believe asking authors to address specific ethics questions in their papers serves two goals: (i) allowing the authors to describe their context-sensitive thinking about the appropriateness of their measurements and (ii) allowing the research community to work towards communal norms.

4. STORED DATA

Data storage has its own ethical considerations. Most problematic in this regard is traces of actual network traffic. The act of collecting such data causes no tangible harm as it is merely observation without any interaction by the researcher. However, the disclosure of confidential information recorded during such measurements can in fact cause serious harm.

Again, there is not a one-size-fits-all approach to managing the associated risk. For instance, even though they both passively record network traffic, there is a vast difference in terms of sensitivity between NetFlow records and full payload packet traces. Also, some work requires actual IP addresses be left intact, whereas other work can be readily conducted with anonymized IP addresses. Therefore, the precautions one takes with each may well be quite different. We do not prescribe a set of techniques but ask researchers to sketch their approach to keeping the data from accidental disclosure.

Also, as the discussion of trends above highlights, it is reasonable to expect that a decade from now, a re-examination of a dataset using newly developed algorithms could extract far more information than is currently possible. Such analysis could therefore present privacy and confidentiality concerns not well understood today. As a practical example, consider that datasets published in the 1980s and early 1990s could probably be mined for passwords using the packet timing algorithms published in 2001[8]. This may or may not be relevant to today depending on whether past password practices of persons remain germane today.

We do not believe it is reasonable to expect researchers to anticipate the results of future research. What is reasonable is to expect a research team to understand how current techniques could mine their measurement data and to safeguard their data against exploitation.

5. QUESTIONS FOR AN ETHICS SECTION

Our position is that every paper on measurement would contain a section on ethical considerations that answered a short list of questions. We envision that the list of questions would be part of each conference's or journal's call for papers. Or the call for papers could incorporate the questions by reference, for instance by pointing a common list of questions developed by the conference's sponsoring professional society.

We strive for a short list of questions. We believe that a short list that captures 80% of the ethics issues is better than a much longer list that is still not exhaustive and would consume large amounts of a paper's page budget. Of course, when thornier ethics issues are present the authors should describe those, as well, even though they are not directly addressing one of the questions posed. As a strawman, we suggest the following questions:

- *Could the collection of the data in this study be reasonably expected to cause potential tangible harm to any person's well-being? If so, discuss measures taken to mitigate the risk of tangible harm.*
- *Using current techniques, can the data collected in this study reveal private or confidential information about individuals? If so, discuss measures taken to keep the data protected from inappropriate disclosure or misuse?*

These questions intentionally do not address several items:

- As discussed in § 3 and § 4, we do not prescribe remedies.
- There is no suggestion of when it might be appropriate to consult an Institutional Review Board (IRB) or similar body. These institutional bodies have their place, but given their institutional nature do not provide a proxy for the community's ethical review.
- We do not attempt to assess the ethics of the research *result*. Rather our focus is on the ethics of the research process. Researchers are committed to advancing knowledge and in our view, that includes publishing results and techniques that may, if used unethically, cause tangible harm.[6]

6. CONCLUSIONS AND FUTURE WORK

This paper presents a strawman suggestion that we ask authors of measurement papers to include a (short) ethics section in their paper. Further, this section is not a nebulous call to "discuss ethics", but focuses on a few specific questions designed to elicit answers to ethics issues that research teams should have considered in the design and execution of their measurement experiments. The questions are not designed to be comprehensive, but rather to surface common issues. While we would be delighted to see the community adopt our suggestion, our goals are more modest. We would like to encourage the research community to find a way to effectively surface ethics questions surrounding individual measurement studies in a way that allows program committees to better evaluate the ethics of a measurement experiment, and allows the broader community to benefit.

[6] In the limit, anything that sources packets can be coaxed into doing harm—even if only by sending at inordinate rates.

7. REFERENCES

[1] *Software Engineering Code of Ethics and Professional Practice; Version 5.2*. 1999.

[2] D. Bertsimas and L. D. Servi. Deducing queueing from transactional data: the queue inference engine revisited. *Operations Research*, 40:217–228, 1992.

[3] D. Cousins, C. Partridge, K. Bongiovani, A. W. Jackson, R. Krishnan, T. Saxena, and W. T. Strayer. Understanding encrypted networks through signal and systems analysis of traffic timing. *Proc. 2003 IEEE Aerospace Conference*, Mar. 2003.

[4] D. Dittrich and E. Kenneally. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. U.S. Department of Homeland Security, Aug. 2012.

[5] G. J. Holzmann and B. Pehrson. *The Early History of Data Networks*. IEEE Computer Society, 1995.

[6] R. C. Larson. The queue inference engine: Deducing queue statistics from transactional data. *Management Science*, 36:586–601, May 1990.

[7] D. C. Sicker, P. Ohm, and D. Grunwald. Legal issues surround monitoring during network research. In *Proc. ACM Internet Measurement Conference*, Oct. 2007.

[8] D. X. Song, D. Wagner, and Z. Tian. Timing analysis of keystrokes and timing attacks on ssh. *USENIX Security*, 2001.