Ethical Concerns for Censorship Measurement

Ben Jones, Roya Ensafi, Nick Feamster, Vern Paxson, Nick Weaver

Princeton University, UC Berkeley, International Computer Science Institute

Abstract

Based on our experiences in measuring censorship in several projects, we frame various ethical questions and challenges that we have encountered. We offer this short document to highlight open questions that we view as important to consider when establishing ethical norms for censorship measurement.

1 Introduction

Various sub-communities in computer science research, including the networking and security research communities, have taken an increasing interest in characterizing information control and censorship in various countries around the world. This research is driven by the desire to have reliable, continuous, and comprehensive empirical data about the nature of censorship around the world. Such data may be valuable to political scientists and sociologists who study government controls, to designers of systems that aim to circumvent these controls, and even to the general public, who may benefit from greater transparency concerning government controls over information.

Despite the value of such data, it is challenging to gather representative measurements about Internet censorship and control—perhaps even more so than conventional Internet measurements. Because Internet censorship and control is both volatile across time and variable by region, obtaining widespread coverage and performing continuous measurements have significant importance. Yet, even for conventional network measurement, obtaining such comprehensive measurements poses challenges; the problem becomes even more vexing when performing these measurements might implicate the owners of the devices that perform the measurements. Thus, in addition to obtaining *coverage*, designers of censorship measurement tools must design tools that preserve the *safety* of humans that own the devices that perform these types of measurements.

To date, there are three general approaches to gathering these types of measurements:

- *Deploy researchers with software.* One approach, commonly employed by organizations such as the Citizen Lab, is to send researchers to countries to directly perform censorship measurements. This approach can gather measurement snapshots, but does not scale well for acquiring continuous measurements (since the researchers do not live in the country), and also potentially places researchers in harm's way.

- *Deploy software to citizens.* Another approach is to entice citizens and activists who already live in the country to install or deploy software that performs measurements. This approach may sometimes achieve more continuous measurements, but it does not always achieve continuity, and it also potentially places people in harm's way.

- *Co-opt existing deployed software.* A third approach is to send traffic from third-party sites towards measurement targets of interest to induce those targets to collect measurements about censorship. Wright *et al.* described this general approach and related ethical concerns, but did not implement any specific method [3]. We have employed this approach in several previously published studies (*e.g.*, Spooky scanning [2], Encore [1]) and are also exploring this approach by way of measurement through open resolvers. This approach achieves continuous, widespread coverage, but faces issues of user consent, since the owners of the targets may be unwittingly implicated for performing such measurements.

Because they ensure that all participants have consented to collecting the measurements, the first two approaches raise fewer ethical questions. Yet, the data they collect may ultimately not suffice for certain measurement goals, since they have neither coverage across time nor fine-grained coverage across regions.

The lack of coverage that plague the first two methods can make it difficult to study certain phenomena of interest. For example, assessing how information control varies during certain events (*e.g.*, an election, a protest) requires having a suitable set of *baseline* measurements for the behavior before and after the event of interest; doing so also requires sufficient measurements to disambiguate overt censorship from transient network problems. These shortcomings bring us to the third method, which offers substantial additional capabilities and benefits, but introduces new risks and ethical considerations, since it involves triggering measurements that might implicate users in performing illicit or illegal measurements without their knowledge.

2 Measurements from Co-Opted Devices

Motivated by the desire to capture more widespread measurements, we have designed several different measurement tools that collect measurements of filtering and censorship through indirect means:

- Spooky scanning, which induces machines to send TCP SYN-ACK and RST packets to possible censorship targets by sending spoofed packets from third-party locations [2].

- Encore, which induces a user's browser to visit a possible censorship target through a mechanism known as a cross-origin request [1].

Our experience with both human subjects review boards (specifically, our universities' IRBs) and the networking and security communities

NS Ethics'15, August 17–21 2015, London, United Kingdom

ACM 978-1-4503-3541-6/15/08$15.00
DOI: http://dx.doi.org/10.1145/2793013.2793015

with this line of work suggest that these mechanisms fall into an ethical grey area. The main ethical quandary is: *To what extent can a user be implicated for traffic that leaves their machine towards a potentially censored destination?*

The first, most basic point to note is that universities and research organizations do not currently have review boards equipped to evaluate these research methods. IRBs evaluate research protocols relating to "human subjects" experiments, which describe a specific type of research involving intervention with people, typically to collect individualized data directly from them. Measurement of the technical specifics of censorship (what content the censor blocks, and technically how they impose the blocking) falls outside of human subjects research, and thus outside the purview of university IRBs. Yet, although the experiments do not involve human subjects, they nonetheless involve *potential risk to people*.

Even more challenging, the actual degree of risk is often very difficult to ascertain. Depending on available resources, researchers might be able to ascertain the legality within different countries of conducting particular measurements. However, for some regimes, legality of method does not necessarily equate to safety for implicated subjects. In addition, these subjects could face privacy hazards (e.g., being falsely implicated in accessing salacious content). Wright *et al.* note that these concerns are especially important when the content being accessed is potentially illegal [3]. Conversely, measurement methods that in some countries technically violate the law might in practice not create any real hazards. (We touch on the ethics of conducting illegal measurements in the next section.)

Of the two types of measurements we outline above, to date the research community appears somewhat more comfortable with Spooky scanning versus Encore. The difference likely arises because Spooky scan's traffic manipulation concerns only layers 3 and 4, rather than the application layer, and has similarities to probing that researchers have performed for other purposes. Related to this, the Princeton University IRB acknowledged that, while they were not equipped to evaluate ethical questions, it seemed unreasonable to expect that any user could control the traffic that one of their devices initiated, since much of this traffic cannot be reasonably traced to human action. They specifically pointed out that a user may have malware or spyware installed; are they to be held personally liable for the traffic that they are unwittingly generating by hosting that malware?

Similarly, are users to be held responsible for the traffic that third-party traffic generates to different sites? While the answer to these questions seems to be "no"—and, indeed, this was the rationale that Princeton offered us—we need to consider that governments may not always proceed with what we technologies view as a reasonable way to analyze activity. Their assessment may incorporate ulterior motives unrelated to the technical specifics, or the parties conducting the assessment may simply lack sufficient "clue" to understand technical nuances that make it clear a given user in fact did not participate in seeming communication. The ethical lines become further blurry when the software is specifically designed to send traffic towards sites that may be blocked or illegal.

The typical approach to performing experiments that pose potential risks to humans is to obtain consent. Even in human subjects experiments, however, it is recognized that consent may not always be required when obtaining it would interfere with the experiment's results (*e.g.*, if the human subjects experiment involves deception, then consent is often not required, since it would tip off the participants to the very phenomenon being studied). IRBs explicitly weigh benefits (both individual, and societal) versus risks in making such decisions.

Consent introduces similar challenging questions for censorship measurement. For one, consent reduces the likelihood of a continuous set of measurements with widespread coverage; obtaining measurements at scale with consent may be impossible. Second, consent itself may place users in more danger than if their devices unwittingly participate in traffic. Another interesting consideration along these lines is that the more widespread co-opted measurements become, the more protection a user receives—for example, the prevalence of malware and third-party trackers itself lends credibility to the argument that a user cannot reasonably control the traffic that their devices send.

3 Questions and Considerations

Ultimately, we have succeeded in publishing some of our measurement methods, but more widespread measurements *using* these techniques remains in the balance. For example, we used Encore to measure Web filtering to Facebook, YouTube, and Twitter, under the argument that nearly all sites have embedded content to these sites already (*e.g.*, Facebook's "thumbs up" button inherently induces the browser to send traffic to Facebook already anyhow). Whether we can use these types of methods and others (*e.g.*, measurements through open DNS resolvers) to collect more widespread measurements hinges on the answers to several important questions and touchstones:

- When is consent necessary? Do the benefits of widespread measurement outweigh the need for consent? Can we mitigate or eliminate the need for consent by *tagging* measurements with some indication that they were collected without the user's consent? For example, we can consider embedding in packet payloads text pointing to a web page explaining the research project. (Different mechanisms for doing so may presume different levels of censor analyst "clue", as discussed below.)

- Does identifying intermediary machines that arguably constitute "infrastructure" (*e.g.*, an ISP's DNS resolver, a CDN cache node), and thus presumably have no direct associations with the actions of individuals, sufficiently mitigate risk? Are there tradeoffs we face by restricting ourselves to only measuring from such infrastructure machines?

- How should we consider community norms and the shades-of-grey that come into play when scanning for services that are open but arguably not meant to be. What makes a service "fair game" to co-opt for measurement?

- What considerations should censorship measurement incorporate to respect the resources that the measurement imposes on third parties, and the time sites spend investigating measurement traffic? Here we can consider techniques such as "tagging" (discussed above) and informative DNS PTR records for associated IP addresses.

- If a given country deems the network traffic associated with censorship measurements itself in violation of the country's laws, on what basis can such measurements ethically proceed? Are some forms of illegal measurements more ethical in this regard than others?

- How do we manage the uncertainty in risk to users due to differing technical abilities and adherence to the rule of law

across countries? Is it reasonable to assume that a censor who analyzes network logs for malfeasance will fully understand (i.e., have "clue") what the logs contain, and take the time and effort to look more broadly than just at specific infringing actions?

- Along with considering risks, what are apt ways to assess the benefits provided by censorship measurements?

We do not presume to have the answers to these questions, but our experience in the design of various censorship tools offers what we believe reflect useful perspectives for dialogs with both the broader community and with ethicists about how to strike the right balance between the substantial benefits that censorship measurements can provide with the unknown (and potentially serious) risks associated with unbridled measurements in this space.

Acknowledgments

Ben Jones is partially funded by the Open Technology Fund's Information Control Fellows Program. This work is also supported by NSF Awards CNS-1540066/CNS-1223717/CNS-1237265, and a Google Focused Research Award. We thank Michael Tschantz for his input on an earlier draft and for helpful discussions that helped formulate our thinking.

References

[1] S. Burnett and N. Feamster. Encore: Lightweight Measurement of Censorship with Cross-Origin Requests. *ACM SIGCOMM*, Aug. 2015. (Cited on page 1.)

[2] R. Ensafi, J. Knockel, G. Alexander, and J. R. C. l. Detecting intentional packet drops on the internet via TCP/IP side channels. In *Passive and Active Measurement Conference*. Springer, 2014. (Cited on page 1.)

[3] J. Wright, T. de Souza, and I. Brown. Fine-grained censorship mapping: Information sources, legality and ethics. In *Free and Open Communications on the Internet*. USENIX, 2011. (Cited on pages 1 and 2.)