Data&Society

No Encore for Encore? Ethical questions for web-based censorship measurement[1]

Case Study 09.14.2015 (Draft)

ARVIND NARAYANAN, BENDERT ZEVENBERGEN

A note to instructors

If you are planning to use this case study in class, we suggest the following order of activities for students: first reading Sections 1 – 3, then discussing the ethical questions that arise (Section 4 has some questions that may help guide this discussion), and finally reading our ethical analysis in Section 5.

[1] Arvind Narayanan is a computer scientist at Princeton. He advised a Master's thesis, described in Section 2, that utilized a similar methodology to the Encore project. Bendert Zevenbergen is a Ph.D candidate and researcher at the Oxford Internet Institute where he studies the intersection of law, ethics, social science, and the Internet. Along with a colleague at OII, he first brought certain ethical concerns to the Encore authors' attention, resulting in a significant change to the design. This case study is the result of a dialogue between us.

This case study was first written for the Council for Big Data, Ethics, and Society. Funding for this Council was provided by the National Science Foundation (#IIS-1413864). For more information on the Council, see: http://bdes.datasociety.net/

We are grateful for useful feedback from Nick Feamster, Jacob Metcalf, Matt Salganik, Joss Wright, and members of the BDES council.

1. Provocation

```
<iframe src="//encore.noise.gatech.edu/task.html"
            width="0" height="0"
            style="display: none">
</iframe>
```

Anyone who administers a web page can copy-paste the above snippet into the source code of the page. It comes from the Encore project at the Network Operations and Internet Security Lab at Princeton, formerly at Georgia Tech. Its effect is to inject an invisible element into the page, which will then instruct the visitor's browser to download and execute a piece of code.[2] The code in question performs *censorship measurement*: it further instructs the visitor's browser to access content from one of various potentially filtered websites — again invisibly — and report back to the research team's server whether or not the access attempt was successful. By aggregating data from visitors to websites that deploy this measurement code snippet and inferring these visitors' locations based on their IP addresses, researchers can obtain an accurate and up-to-date view into web filtering worldwide.

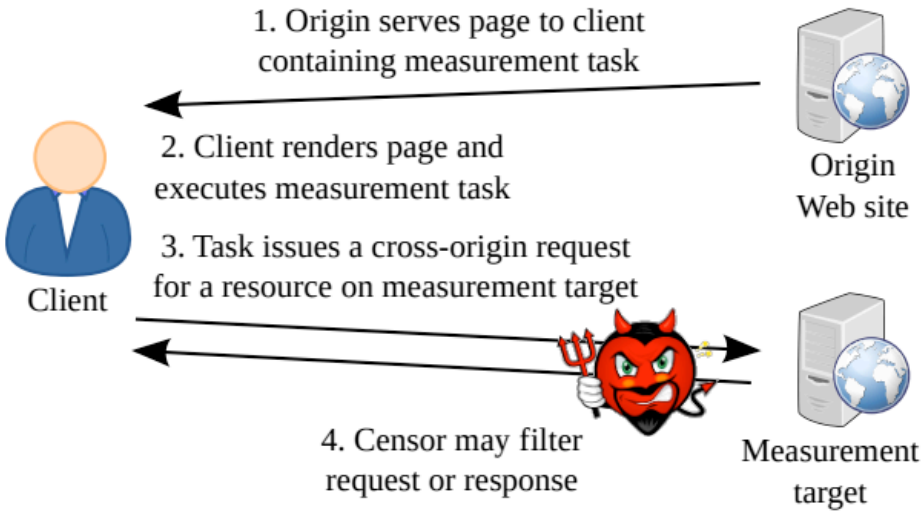

Figure: How Encore works. Reproduced from Burnett and Feamster's paper.[3]

[2] The code is located at http://encore.noise.gatech.edu/task.html and can be viewed using your web browser's source code viewing feature.
[3] Burnett, Sam, and Nick Feamster. "Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests," *SIGCOMM '15*, August 17–21, 2015, London, United Kingdom, accessed August 21, 2015, http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p653.pdf.

The researchers, Sam Burnett and Nick Feamster, used this technique to conduct measurements for a period of seven months, as of January 2015, via installations by at least 17 volunteers. Measurements were recorded from 88,260 distinct IP addresses in 170 countries, with China, India, the United Kingdom, and Brazil reporting at least 1,000 measurements, and more than 100 measurements from Egypt, South Korea, Iran, Pakistan, Turkey, and Saudi Arabia.

Encore revealed valuable information about the censorship activities of these governments, but it did so by altering the behavior of computers in ways that users were probably not anticipating and had not consented to. Encore is thus one example of a growing ethical conundrum for computer science and computer security researchers: should researchers be permitted to surreptitiously alter the behavior of Internet-connected devices in order to gain scientific data about the behavior of users and networks? If they should be allowed to in at least some cases, what are the criteria for determining proper and improper uses of these techniques and who should enforce such standards? Encore also throws into sharp relief the conflict between two objectives: building automated, large-scale, globally applicable measurement tools and carefully analyzing ethical issues with consideration to all relevant stakeholders, laws, norms, and social, cultural, and political contexts.

2. Technical background

The architecture of the Internet, and the web in particular, affords a variety of ways of observing the behavior of networked devices on a large scale without the cooperation of users. Indeed, the multi-billion-dollar online ad targeting industry is built on this idea. Invisible "third parties" track our devices as we browse the web to build profiles of our interests and behavior; the average top-50 website contains 64 tracking mechanisms.[4] Meanwhile, analytics firms track people in physical spaces like shopping malls based on the WiFi, cellular, and other emanations from their smartphones.[5]

Computer science researchers have also made creative use of methods that allow observing devices without affirmative user consent. These studies have led to insights on the state of computer security, the economics of online advertising and of spam campaigns, censorship and filtering around the

[4] Angwin, Julia. "The Web's New Gold Mine: Your Secrets." *The Wall Street Journal*, July 30, 2010, accessed August 11, 2015, http://www.wsj.com/articles/SB10001424052748703940904575395073512989404.
[5] Euclid Analytics website, accessed August 11, 2015, http://euclidanalytics.com/.

world, and more.

The most intrusive of these studies, technologically speaking, are those that exploit unpatched security flaws to turn users' devices into observation points. A well-known one is *Spamalytics,* a study where researchers took over control of a botnet — a network of machines infected with malware and controlled by a single operator — to modify and study the spam campaigns that originated from the infected machines.[6] In another instance, an anonymous researcher or researchers *created* a botnet named Carna by infiltrating over 400,000 devices such as routers whose default passwords hadn't been changed, and used it to study essentially the entirety of Internet-connected devices.[7] [8]

Other studies are non-intrusive: they simply eavesdrop on network traffic without interfering with devices. In computer networking, analyzing traffic data for improving performance and testing new protocols is standard practice, and arguably essential. Such studies typically make use of data provided by Internet Service Providers (ISPs) that can be staggeringly large in size. For example, a 2014 study of IPv6 adoption utilized (among others) a dataset of traffic statistics that covered an estimated 33–50% of all internet traffic for 2013.[9] This type of research generally looks at network traffic in the aggregate rather than the behavior of individual users or devices, but there are exceptions. One study used traffic metadata of millions of users, including a campus network, to study the economics of online advertising.[10] They did this by inferring what information is collected about individuals by advertisers as well as how expensive the ads being shown to them were, and analysing the relationship between the two.

Peer-to-peer networks are particularly amenable to non-intrusive study. Since such networks are designed to route information among peers rather than to and from designated servers, a researcher can simply set up one or more peers and hop on board, without needing any special privileges such as

[6] Kanich, Chris, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage. "Spamalytics: An Empirical Analysis of Spam Marketing Conversion," *CCS'08*, October 27-31, 2008, Alexandria, VA, accessed August 11, 2015, http://www.icsi.berkeley.edu/pubs/networking/2008-ccs-spamalytics.pdf.
[7] "Internet Census 2012: Port scanning /0 using insecure embedded devices," Carna Botnet, accessed August 11, 2015, http://internetcensus2012.bitbucket.org/paper.html.
[8] Krenc, Thomas, Oliver Hohlfeld, and Anja Feldmann. "An Internet Census Taken by an Illegal Botnet – A Qualitative Assessment of Published Measurements," *ACM SIGCOMM Computer Communication Review*, Vol. 44, No. 3, July 2014, accessed August 11, 2015, http://www.net.t-labs.tu-berlin.de/papers/KHF-ICIBQ-14.pdf.
[9] Czyz, Jakub, Mark Allman, Jing Zhang, Scott Iekel-Johnson, Eric Osterweil, and Michael Bailey. "Measuring IPv6 Adoption," *SIGCOMM'14*, August 17–22, 2014, Chicago, IL, accessed August 11, 2015, http://nsrg.eecs.umich.edu/publications/sigcomm14_ipv6.pdf.
[10] Gill, Phillipa, Vijay Erramilli, Augustin Chaintreau, Bala Krishnamurthy, Dina Papagiannaki, and Pablo Rodriguez, "Follow the Money: Understanding Economics of Online Aggregation and Advertising," *IMC'13*, October 23–25, 2013, Barcelona, Spain, accessed August 11, 2015, http://conferences.sigcomm.org/imc/2013/papers/imc184s-gillAemb.pdf.

co-operation from an ISP. The BitTorrent file-sharing system, the Tor anonymity network, and the Bitcoin cryptocurrency network have all been studied this way.[11] [12] [13]

A burgeoning category of research lies in between these two in terms of intrusiveness: methods that use active probing of devices in some way but not exploitation of any security holes. These techniques are both technically and ethically fascinating, and include the Encore study.

An archetypal example of active probing is *network scanning* for security assessment of networks[14]; the ZMap research tool allows performing fast scans on an Internet-wide scale.[15] Network scanning has a long history, but a variety of new techniques are stealthier. *Idle port scanning* uses "side-channel attacks" to bounce traffic off an Internet-connected device to make measurements of *other* devices.[16] These side-channel attacks are different from exploitation of security bugs: the researcher doesn't take control over devices in any way; the bugs that allow side channels are common to many different implementations, and may be inherent to the protocol specification.

Encore similarly makes use of unintended effects that are inherent in the architecture of web rather than a bug in any specific browser. The "same-origin policy" is intended to quarantine content from different domains even when they are loaded side-by-side on the same page, but there are limits to the effectiveness of this protection. Recent research at Princeton created an interesting twist to Encore's research methodology, showing how to deploy these measurements through online advertisements.[17] The researcher simply purchases ad impressions — available cheaply by the thousands — and delivers the measurement code as part of the ad. This technique allows targeting by geography and

[11] J.A. Pouwelse, P. Garbacki, D.H.J Epema, and H.J. Sips, "The Bittorrent P2P File-Sharing System: Measurements and Analysis," Department of Computer Science, Delft University of Technology, the Netherlands, accessed August 11, 2015, http://www.cs.unibo.it/babaoglu/courses/cas04-05/papers/bittorrent.pdf.
[12] Winter, Philipp, Richard Köwer, Martin Mulazzani, Markus Huber, Sebastian Schrittwieser, Stefan Lindskog, and Edgar Weippl, "Spoiled Onions: Exposing Malicious Tor Exit Relays," Karlstad University, Sweden, SBA Research, Austria, and FH Campus Wien, Austria, accessed August 11, 2015, http://www.cs.kau.se/philwint/spoiled_onions/pets2014.pdf.
[13] Miller, Andrew, James Litton, Andrew Pachulski, Neal Gupta, Dave Levin, Neil Spring, and Bobby Bhattacharjee. "Discovering Bitcoin's Public Topology and Influential Nodes," University of Maryland, College Park, accessed August 11, 2015, https://cs.umd.edu/projects/coinscope/coinscope.pdf.
[14] Nmap ("Network Mapper"), accessed August 11, 2015, https://nmap.org/.
[15] Durumeric, Zakir, Eric Wustrow, and J. Alex Halderman. "ZMap: Fast Internet-Wide Scanning and its Security Applications," Proceedings of the 22nd USENIX Security Symposium, August 2013, accessed August 11, 2015, https://zmap.io/paper.pdf.
[16] Ensafi, Roya, Jong Chun Park, Deepak Kapur, and Jedidiah R. Crandall. "Idle Port Scanning and Non-interference Analysis of Network Protocol Stacks Using Model Checking," Department of Computer Science, University of Mexico, accessed August 11, 2015. https://www.usenix.org/legacy/event/sec10/tech/full_papers/Ensafi.pdf.
[17] Zimmerman, Peter Thomas. "Measuring Privacy, Security, and Censorship Through the Utilization of Online Advertising Exchanges," A Thesis Presented to the Faculty of Princeton University in Candidacy for the Degree of Master of Science in Engineering, June 2015, accessed August 11, 2015, ftp://ftp.cs.princeton.edu/techreports/2015/988.pdf.

demographics and can reach any user without relying on deployment by a website that's reachable by the user.

Encore is part of the small but growing research area of censorship measurement that sees a handful of significant publications each year. In the United States, funding for censorship measurement comes from the National Science Foundation, indirectly from the State Department through its funding of censorship circumvention research, and from a few companies and philanthropic organizations.

The most basic objective of censorship measurement is compiling data on what is censored or filtered, when, and for which users. A prominent example is the Harvard Berkman Center's Herdict project that aims to crowd-source and aggregate data about web filtering.[18] The Tor project's Open Observatory of Network Interference (OONI) has a similar aim, but provides a downloadable script that users can run.[19] Such data collection is sometimes straightforward but sometimes requires technical innovation and research. Encore, of course, is one example; another is ConceptDopplr, a tool that incorporates a way to efficiently probe a keyword-based blocking system to discover the set of all blacklisted keywords.[20]

Another objective of censorship measurement is to understand the technical mechanisms by which censorship operates. Here are some questions on which researchers have been able to shed light: do governments operate filters in a centralized way at Internet routers at the nation's borders, or in a decentralized way closer to the users?[21] How quickly are censors able to remove content from microblogging sites?[22] Does censorship operate purely by blocking or removal of content, or are performance degradation and modification of content also part of the picture?[23] Do censors have the

[18] Herdict, Berkman Center for Internet & Society at Harvard University, accessed August 11, 2015, https://www.herdict.org/.
[19] Filasto, Arturo, and Jacob Appelbaum. "OONI : Open Observatory of Network Interference," The Tor Project and the University of Washington, accessed August 11, 2015, https://www.usenix.org/system/files/conference/foci12/foci12-final12.pdf.
[20] Crandall, Jedidiah R., Daniel Zinn, Michael Byrd, Earl Barr, and Rich East. "ConceptDoppler: A Weather Tracker for Internet Censorship," paper for the 14th ACM Conference on Computer and Communications Security, October 29 - November 2, 2007, accessed August 11, 2015, https://www.cs.unm.edu/~crandall/concept_doppler_ccs07.pdf.
[21] Xu, Xueyang, Z. Morley Mao, and J. Alex Halderman. "Internet Censorship in China: Where Does the Filtering Occur?" Department of Computer Science and Engineering, University of Michigan, accessed August 11, 2015, https://web.eecs.umich.edu/~zmao/Papers/china-censorship-pam11.pdf.
[22] Zhu, Tao, David Phipps, Adam Pridgen, Jedidiah R. Crandall, and Dan S. Wallach. "The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions," paper for the 22nd USENIX Security Symposium in Washington, DC, August 2013, accessed August 11, 2015, http://arxiv.org/ftp/arxiv/papers/1303/1303.0597.pdf.
[23] Burnett, Sam, and Nick Feamster. "Making Sense of Internet Censorship: A New Frontier for Internet Measurement," *ACM SIGCOMM Computer Communication Review*, Vol. 43, No. 3, July 2013, accessed August 11, 2015, http://www.sigcomm.org/sites/default/files/ccr/papers/2013/July/2500098-2500111.pdf.

technical infrastructure to examine the entire contents of internet traffic ("deep packet inspection") without slowing it down, or do they only look at the metadata?[24] What sorts of collateral damage does censorship cause?[25] So far, the bulk of the computer science research on censorship measurement has been devoted to this class of questions.

Moving from the realm of computers and networks to the realm of people and governments, political scientists are interested in internet censorship in terms of the motives of censors, the impact of censorship on freedom of speech, and so on.[26] Measurement directly or indirectly helps answer these questions. In 2013 Harvard researchers analyzed millions of social media posts to show that censorship in China allows government criticism but silences collective expression.[27]

For obvious reasons, it's hard to measure censorship from a vantage point outside the country or countries of interest. There are some interesting and important exceptions. When censorship of online posts happens after the posts are published *and* if researchers are able to get to the content before the censors do, measurement can happen from the outside.[28] In another instance, a hacktivist group leaked 600 gigabytes of log files of internet filtering devices used in Syria, allowing researchers to gain insights into censorship in that country.[29] Similarly, an anonymous ISP in Pakistan provided researchers access to a trove of data that enabled analysis of Pakistani censorship.[30]

Outside such exceptions, collecting data about censorship requires the participation of volunteers "on the ground" — volunteers who might expose themselves to some risk and whose number limits the scale of measurements. Encore straddles these two categories: it avoids the need for researchers to recruit volunteers and is easily scalable, but the individuals whose devices are used do face potential

[24] Wilde, Tim, "Knock Knock Knockin' on Bridges' Doors," *The Tor Project*, January 7, 2012, accessed August 11, 2015, https://blog.torproject.org/blog/knock-knock-knockin-bridges-doors.
[25] Dainotti, Alberto, Claudio Squarcella, Emile Aben, Kimberly C. Claffy, Marco Chiesa, Michele Russo, and Antonio Pescape´, "Analysis of Country-wide Internet Outages Caused by Censorship," *2013 IEEE*, accessed August 11, 2015, http://www.caida.org/publications/papers/2014/outages_censorship/outages_censorship.pdf.
[26] Morozov, Evgeny. *The Net Delusion: The Dark Side of Internet Freedom*. New York: PublicAffairs, 2012. Print. http://www.amazon.com/The-Net-Delusion-Internet-Freedom/dp/1610391063.
[27] King, Gary, Jennifer Pan, and Margaret E. Roberts. "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, May 2013, accessed August 11, 2015, http://gking.harvard.edu/files/censored.pdf.
[28] Ibid.
[29] Chaabane, Abdelberi, Terence Chen, Mathieu Cunche, Emiliano De Cristofaro, Arik Friedman, and Mohamed Ali Kaafar. "Censorship in the Wild: Analyzing Internet Filtering in Syria," *IMC'14*, November 5–7, 2014, Vancouver, BC, Canada, accessed August 11, 2015, http://conferences.sigcomm.org/imc/2014/papers/p285.pdf.
[30] Khattak, Sheharbano, Mobin Javed, Syed Ali Khayam, Zartash Afzal Uzmi, and Vern Paxson. "A Look at the Consequences of Internet Censorship Through an ISP Lens," *IMC'14*, November 5–7, 2014, Vancouver, BC, Canada, accessed August 11, 2015, http://conferences.sigcomm.org/imc/2014/papers/p271.pdf.

risks. Encore also provides a geographically fine-grained view of measurement, which researchers in the field value.[31] Further, since Encore turns regular Internet users into measurement vantage points, it avoids the problem of censors being able to detect and disable measurement units.

3. The Program Committee's reaction

Which of the research projects we've looked at should be considered human-subjects research? Questions of this sort have long been contentious in computer science. Human-subjects research at institutions that receive federal funding in the United States is subject to Institutional Review Board (IRB) oversight. IRBs approve research proposals based on efforts to account for and mitigate risk to participants; typically research that poses little risk to individual human subjects is categorized as "exempt" from extensive oversight. In practice, however, much of such computer science research has operated without IRB involvement. Historically computer science and engineering has considered itself to be researching human-less systems, and university IRBs are typically geared toward regulating biomedical and social science research. When IRBs encounter computer science research there is often mutual confusion.

Acknowledging that there are ethical concerns whether or not human subjects are involved, researchers have sought an alternative way to ensure that published research is justifiable on scientific ethics grounds. Several sub-communities including computer security, networking, and Internet measurement — which collectively encompass all of the research described above — appear to be converging on conference program committees as the oversight mechanism. What's a program committee? The most prestigious research in computer science is published in the proceedings of conferences rather than journals; each iteration of each conference selects a program committee to carry out peer review.

The appropriateness of ethical gatekeeping by program committees is a topic of continual debate in the

[31] Wright, Joss, Tulio de Souza, and Ian Brown. "Fine-Grained Censorship Mapping Information Sources, Legality and Ethics," Oxford Internet Institute and Oxford University Computing Laboratory, accessed August 11, 2015, http://static.usenix.org/event/foci11/tech/final_files/Wright.pdf.

community [see, for example[32]]. Supporters of this model argue that technical domain expertise is critical for ethical review and that each subcommunity must evolve its own norms by adapting ethical principles to the specific domain.

On the other hand, the system has numerous shortcomings, some inherent and others potentially fixable. First, the review happens after the research is complete. Unlike IRBs, there is no process for advance or continual review. The uncertainty induced by the potential rejection of research by program committees might lead to researchers abandoning some research ideas or entire areas of research — especially research that pursues methodological innovation — even if, on balance, the research would have been found to be ethically acceptable if it had proceeded. In cases where the putative harm arises from *conducting* the research rather than its publication, the retrospective ethical review in fact fails to prevent that harm.

Second, program committees are composed of domain experts and rarely include any members with scholarly expertise in research ethics or ethics in general. Third, since they are formed and disbanded for each conference, they have essentially no capacity for institutional memory — whether about specific research projects or about decision-making critieria and procedures. Unsurprisingly, they have operated, at least so far, without consistency in ethical standards and with ad-hoc decision-making processes; indeed, to our knowledge, not a single one has published rules or guidelines for what qualifies as ethical research as part of the call for papers!

The Encore paper was submitted to ACM SIGCOMM 2015, a prestigious conference on computer networking. After heated debate, the committee accepted the paper for publication, but with a "signing statement" at the top of the paper, an unprecedented move.[33]

The committee's ethical objections stemmed from several arguments, outlined in the public review of the paper.[34] First, third-party requests used for ad tracking, the committee held, at least notionally reflect the user's intent, whereas Encore's requests do not. Second, users downloading censored URLs

[32] Kenneally, Erin, and Michael Bailey."Cyber-security Research Ethics Dialogue & Strategy Workshop," *ACM SIGCOMM Computer Communication Review*, Vol. 44, No. 2, April 2014, accessed August 22, 2015, http://mdbailey.ece.illinois.edu/publications/ccr-2014.pdf.
[33] SIGCOMM's call for papers had a requirement (for the first time in 2015) that authors "attest that their work complies with all applicable ethical standards of their home institution(s), including ... policies on experiments involving humans". The actual ethical review went well beyond checking for such compliance.
[34] Byers, John W. "Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests – Public Review," Department of Computer Science, Boston University, accessed August 11, 2015, http://conferences.sigcomm.org/sigcomm/2015/pdf/reviews/226pr.pdf.

might face repercussions if they live in a regime without due process. Third, the committee believed that most users for whom censorship is an issue would be unlikely to consent to Encore's measurements.

4. Discussion questions

1. Encore makes use of web functionality in a creative way to achieve its aims, just as as many advertising networks do. Should credentialed researchers at universities be held to stricter ethical standards because they are producing generalizable knowledge rather than serving advertisements?

2. As discussed above, computer scientists and data scientists use a largely unique method for ethical self-regulation by assigning conference program committees to be ethics gate keepers. What do you see as potential benefits and risks for this method? Did the committee make the right decision in the Encore case?

3. Encore, like much of network measurement research, uses a single automated technical tool to conduct measurements on a vast, global scale. Does this paradigm complicate the researcher's responsibility to act ethically?

4. Encore researchers claimed that it was infeasible to accurately measure potential risk or get informed consent from every subject. For any given research project, if it is infeasible to meet these basic benchmarks of research ethics, is it appropriate to pursue that project? If that caveat applies to an entire field, should that model of knowledge production be abandoned?

5. Given Encore's goals, how would you design it to minimize risks to unsuspecting users and to be as transparent as possible? To what extent does this require a trade-off with Encore's research objectives?

6. Would you be troubled to learn that Encore conducted measurements from your web browser? How do you think your counterpart in Iran or China would answer this question?

7. Does your approval or disapproval of the Encore project hinge on their goal of illuminating Internet censorship? If their purpose was less praise-worthy would you feel different about their research methods?

5. Ethical analysis

There are several analytical frameworks for scientific ethics and regulation. For example, the Belmont Report[35] concerns scientific and medical research involving human subjects, and established "respect for persons," "beneficence" and "justice" as the guiding principles of research ethics. The Common Rule[36] is the federal regulation that tasks IRBs with reviewing research to ensure it meets those principles. The Menlo Report[37] builds on the Belmont Report and translates the scientific ethics research principles into the computer science and network engineering domain. There are many other frameworks and guidelines for this type of research.[38] In our analysis here we'll roughly follow the Menlo report in terms of structure and the set of principles used.

5.1 Who are the stakeholders?

Any Internet user worldwide can stumble upon the invisible Encore script and carry out a censorship measurement. When an unsuspecting Internet user's browser sends a request to a potentially censored website, as instructed by the Encore code, their IP address may be recorded by the server hosting that website, as well as by many intermediaries and potentially unknown third parties. Most significantly, government-mandated censorship systems may also record and try to identify persons who access a censored website, although this is an assumption and may vary significantly by country. The Encore

[35] "The Belmont Report," April 18, 1979, U.S. Department of Health and Human Services, accessed August 11, 2015, http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.
[36] "Federal Policy for the Protection of Human Subjects ('Common Rule')," U.S. Department of Health and Human Services, accessed August 11, 2015, http://www.hhs.gov/ohrp/humansubjects/commonrule/.
[37] "The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research," U.S. Department of Homeland Security Science and Technology Directorate, Cyber Security Division, August 2012, accessed August 11, 2015, https://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/menlo_report_actual_formatted.pdf.
[38] See, for example: Markham, Annette and Elizabeth Buchanan. "Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Committee," approved by the AOIR general membership, December 2012, accessed August 11, 2015, http://aoir.org/reports/ethics2.pdf; and Zevenbergen, Bendert, Ian Brown, Joss Wright, and David Erdos. "Ethical Privacy Guidelines for Mobile Connectivity Measurements," Oxford Internet Institute, University of Oxford, November 2013, accessed August 11, 2015, http://www.oii.ox.ac.uk/research/Ethical_Privacy_Guidelines_for_Mobile_Connectivity_Measurements.pdf.

research team also records the measurements, which includes the user's IP address.

Trying to identify the stakeholders immediately reveals a conflict. Ethics guidelines typically recommend something akin to the following from the Menlo report: "it is first necessary to perform a systematic and comprehensive stakeholder analysis." Yet the worldwide scale of Encore means that analyzing all potential stakeholders individually is infeasible. Worse, the principle is inherently at odds with the goal of *scalability* in computer science and engineering. Scalability is a goal that the Encore authors emphasize; in this context, it means that the team is able to expand the set of measurement targets by simply adding more machine resources and without multiplying the researcher effort required.[39] The dogma of computer science (and the technology industry) enshrines scalability as a virtue. Web companies regularly advertise their ratio of users to engineers, which can be over a million.[40]

The Menlo report briefly acknowledges the issue, stating: "Even a simple link traffic characterization study could involve millions of computers used by humans who are not themselves the direct subjects of research." This tension is a theme to which we will repeatedly return.

Is Encore human-subjects research?

Unsuspecting Internet users across the globe generate research data for the Encore project. Does the reliance on these humans mean that the Encore project constitutes human-subjects research in the traditional sense, analogous to fields such as medical research or psychology? While networking researchers typically see themselves as conducting research on technical systems, the Internet is more properly understood as a sociotechnical system in which humans and technology interact. Experiments on the Internet will likely also include data collection about human behavior, or affect their environment.

Neither the Princeton nor the Georgia Tech IRB considered Encore to be human-subjects research.

[39] There are other, related, meanings of the word scalability. In particular, it can refer to the ability of a computer system to handle greater and greater loads by adding proportionally many computing units in parallel. This sense of the term is simply a sound engineering principle.
[40] Bosworth, Andrew. "Facebook Engineering Bootcamp," Facebook Engineering page, November 19, 2009, accessed August 11, 2015. https://www.facebook.com/notes/facebook-engineering/facebook-engineering-bootcamp/177577963919.

Under the operational definition from the Common Rule that IRBs use, a human subject is a living individual about whom an investigator obtains "(1) Data through intervention or interaction with the individual, or (2) Identifiable private information." Should Encore's collection of IP addresses classify it as human-subjects research?

The question of whether or not IP addresses constitute PII is a well-worn debate.[41] Buchanan et al. note:

> "The Office for Human Research Protections has not issued a formal statement on whether IP addresses are considered to be personally identifiable information for purposes of the HHS protection of human subjects regulations at 45 CFR Part 46. However, for purposes of the HIPAA Privacy Rule, the HHS Office for Civil Rights has opined that an IP address is considered to be a direct identifier of an individual. Other European data regulations consider IPs as identifiers, and as such fall under the realm of the EU Data Directives (1995, 2006). This presents a challenge for international research and should be considered carefully by researchers and boards."[42]

Can Encore's data collection be designed such that the IP addresses collected are generalized so that they are no longer personally identifying, yet the researchers are able to carry out their measurement and analysis objectives? This is an open question.

Under a narrow interpretation, the data that Encore collects is not about the individual but rather about the behavior of censorship systems. On the other hand, the definition stems from medical and behavioral research, and probably did not anticipate the investigator's actions causing *other* parties — in Encore's case, the censor — to collect data about the individual. The Menlo Report advises the investigator to "Respect individuals who are not targets of research yet are impacted" and says that "human subject research should now be considered as "human-harming research" — so the internet users may not be subjects per se, but they can still experience harm due to the research being

[41] McIntyre, Joshua J. "Balancing Expectations of Online Privacy: Why Internet Protocol (IP) Addresses Should Be Protected as Personally Identifiable Information," *DePaul Law Review*, Vol. 60, Issue 3, Spring 2011, accessed August 11, 2015, http://via.library.depaul.edu/cgi/viewcontent.cgi?article=1151&context=law-review.

[42] Buchanan, Elizabeth, John Aycock, Scott Dexter, David Dittrich, and Erin Hvizdak. "Computer Science Security Research and Human Subjects: Emerging Considerations for Research Ethics Boards," *Journal of Empirical Research on Human Research Ethics: An International Journal*, Vol. 6, No. 2, June 2011, pp. 71-83, accessed August 11, 2015, http://www.orau.gov/communityirb/Resources/EmergingConsiderationsForResearchEthicsBoards.pdf.

conducted."

Garfinkel is a proponent of the view that much of computer security research should be viewed as human-subjects research.[43] He proposes what he calls the human test: "would the experiment be useful if the data were generated by a random process and not by a human?" It is not obvious how to apply this test to Encore. If it were practically possible — which, unfortunately, it is not — to replace the humans whose devices are used by Encore for measurement by robots that visit websites in a random fashion, Encore would work very well, and in some ways better than it currently does since biases in measurement times and so on would be minimized.

5.2 Beneficence

The principle of Beneficence concerns the goal of the welfare of research participants and the balancing of probable harms. Ideally, this would require a systematic identification of the probability and magnitude of risk as well as benefits for the stakeholders, a subsequent iterative analysis of minimizing risk and maximizing benefits through the research design, as well as plans to mitigate identified risks and any unforeseen harms that materialize.

Identifying Potential Benefits and Harms

Due to Encore's global scale, it will be tough for a small research group to adhere fully to the requirements of Beneficence. For example, before risks and harms can be identified, they must first be defined. However, due to the complex, dynamic, and innovative nature of the Internet, it is difficult to define the harms for each Internet user concretely, or even for regional groups of Internet users. The norms and attitudes of identified stakeholders with regards to accessing censored content differ greatly around the world, as well as the type of content that is censored, or the enforcement actions that can be triggered. These are influenced by political, religious, historical and other social factors and are difficult — if not impossible — to quantify into a solid assessment of risks for each individual user.

[43] Garfinkel, Simson L. "IRBs and Security Research: Myths, Facts and Mission Creep," Naval Postgraduate School & Harvard University, April 7, 2008, accessed August 11, 2015, https://calhoun.nps.edu/bitstream/handle/10945/40330/garfinkel_IRBs_and_Security_Research.pdf?sequence=1.

Benefits of the research

Internet censorship measurement researchers argue that "whilst filtering and censorship can, to an extent, be open and transparent, their nature tends towards secrecy."[44] Measurement helps illuminate censorship — both its motivations and the technologies behind it. Understanding the motivations behind censorship yield valuable insights in political science, such as the fact that censorship in China allows government criticism but silences collective expression that may spur collective action.[45] Illuminating censorship techniques enhances the ability to create effective censorship circumvention tools.[46]

A view of Internet censorship as harmful to citizens subjected to it is implicit in much of censorship measurement research. Censorship is seen as being in opposition to human rights — the freedom of speech and more specifically the freedom to seek, receive, and impart information.[47]

Computer scientists and engineers also have technical concerns: architecting networks to allow censorship and filtering by governments and intermediaries violates the "end-to-end" principle, a key design philosophy of the Internet. As far back as the year 2,000, Clark and Blumenthal warned that as the end-to-end design erodes, the "Internet might lose some of its key features, in particular its ability to support new and unanticipated applications."[48] Internet engineers also raised this concern, among others, in response to the Stop Online Piracy Act and PROTECT IP Act bills in the United States that proposed to allow the government to block copyright-infringing websites: "Censorship of Internet infrastructure will inevitably cause network errors and security problems. This is true in China, Iran

[44] Wright, Joss, Tulio de Souza, and Ian Brown. "Fine-Grained Censorship Mapping Information Sources, Legality and Ethics," Oxford Internet Institute and Oxford University Computing Laboratory, accessed August 11, 2015, http://static.usenix.org/event/foci11/tech/final_files/Wright.pdf.
[45] King, Gary, Jennifer Pan, and Margaret E. Roberts. "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, 107(02): 326–343, 2013, accessed February 19, 2015, http://core.ac.uk/download/pdf/28941335.pdf.
[46] The Encore authors say, "Researchers, activists, and citizens aim to understand what, where, when, and how governments and organizations implement Internet censorship. This knowledge can [...] guide the development of new circumvention techniques." Burnett, Sam, and Nick Feamster. "Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests," arXiv:1410.1211v2 [cs.NI] July 19, 2015, accessed August 11, 2015, http://arxiv.org/pdf/1410.1211v2.pdf.
[47] "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue," United Nations General Assembly, Human Rights Council, seventeenth session, May 16, 2011, http://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf.
[48] Clark, David D., and Marjory S. Blumenthal. "Rethinking the design of the Internet: The end to end arguments vs. the brave new world," Version for TPRC submission, August 10, 2000, accessed August 11, 2015, http://dspace.mit.edu/bitstream/handle/1721.1/1519/TPRC_Clark_Blumenthal.pdf?sequence=1&origin=publication_detail.

and other countries that censor the network today; it will be just as true of American censorship."[49]

An unequivocally negative view of censorship is not universally held. Bambauer argues that "widespread censorship on-line is not necessarily bad" and that the legitimacy of censorship should be assessed not by what is blocked but rather the transparency and accountability of decision-making regarding censorship.[50] Chu and Cheng view the "Western" lens of individual autonomy and equality as inappropriate for Chinese society and evaluate Chinese online censorship from the perspective of "raising the moral level of both the state and society" [[51] Chapter 1: "Cultural convulsions: examining the Chineseness of cyber China"].

At any rate, scholars have raised critical questions about data science, especially "big data"[52], that apply to censorship measurement as well. For example, since some types of censorship are much easier to detect than others, measurement results may produce a biased picture of the state of censorship and the direction it is moving in. Moreover, stripped of cultural and political context, data is hard to interpret. For instance, it may be easy to find correlations between Internet filtering and news events, but causal attribution is far trickier. Similar concerns have been raised in the field of international development.[53] Censorship measurement researchers should be aware of this debate.

To conclude the discussion of benefits let's recall the principle of Justice, which entails that burdens as well as benefits be fairly and equitably distributed. For example, participants in a study who run the risk of harm should also benefit in the longer run from the research findings. From the Menlo report: "Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit".

Harm: does Encore present more than minimal risk?

[49] Cerf, Vint, et. al. "An Open Letter From Internet Engineers to the United States Congress," December 15, 2011, accessed August 11, 2015, https://www.eff.org/files/internet-engineers-letter.pdf.
[50] Bambauer, Derek E. "Censorship V3.1," *18 IEEE Internet Computing 26* (May/June 2013), Arizona Legal Studies Discussion Paper No. 12-28, September 9, 2012, accessed August 11, 2015, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2144004&download=yes.
[51] Herold, David Kurt, and Peter Marolt (eds.) "Online Society in China: Creating, celebrating, and instrumentalizing the online carnival," Routledge, 2011, accessed August 11, 2015, http://www.tandf.net/books/details/9780415838221/.
[52] boyd, danah, and Kate Crawford. "Critical Questions for Big Data," *Information, Communication & Society*, 15:5, 662-679, (2012) DOI: 10.1080/1369118X.2012.678878, accessed August 11, 2015, http://www.tandfonline.com/doi/pdf/10.1080/1369118x.2012.678878.
[53] Taylor, Linnet, and Dennis Broeders. "In the name of Development: Power, profit and the datafication of the global South," *Geoforum*, Vol. 64, August 2015, pp. 229-237, accessed August 11, 2015, http://www.sciencedirect.com/science/article/pii/S0016718515001761.

The Encore paper itself considers harm primarily in terms of a comparison between Encore usage and regular web browsing. This type of comparison is captured in the notion of "minimal risk" as defined by the Common Rule[54]: "minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research is not greater in and of itself than those encountered during daily life or during the performance of routine physical and psychological examinations or tests".[55]

The authors argue that normal web browsing exposes users to the same risks that Encore does, saying "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" and "laws against accessing filtered content vary from country to country, and may be effectively unenforceable given the ease with which sites (like Encore) can request cross-origin resources without consent."

It is true that the average web user today is not in position to effectively control third-party requests that their browser makes. Tracking technologies often go to great lengths to be stealthy and publishers are often oblivious to the tracking technologies deployed on their properties.[56] Furthermore, online trackers make requests to yet other third parties, just as Encore does. In fact, these "chains" of trackers can be half-a-dozen (or more) deep.[57] While these trackers may not necessarily make requests to censored domains, they beget other risks such as exposing users to surveillance agencies that monitor Internet traffic.[58] [59] Finally, advertisements themselves — and not just advertising networks — can make cross-origin requests to arbitrary domains. The bar to serving ads is much lower than the bar to becoming an advertising network.

However, there are several caveats to this argument and nuances that we should note. First, current

[54] "Proposed Revisions to the Common Rule for the Protection of Human Subjects in the Behavioral and Social Sciences," *National Academy of Sciences*, 2014, accessed August 11, 2015, http://www.ncbi.nlm.nih.gov/books/NBK217976/.
[55] "Code of Federal Regulations," U.S. Department of Health and Human Services, January 15, 2009, accessed August 11, 2015, http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.
[56] Angwin, Julia. "Meet the Online Tracking Device That is Virtually Impossible to Block," *ProPublica*, July 21, 2014, accessed August 11, 2015, https://www.propublica.org/article/meet-the-online-tracking-device-that-is-virtually-impossible-to-block.
[57] Li, Zhou, Kehuan Zhang, Yinglian Xie, Fang Yu, and XiaoFeng Wang, "Knowing Your Enemy: Understanding and Detecting Malicious Web Advertising," *CCS'12*, October 16–18, 2012, Raleigh, North Carolina, accessed August 11, 2015, http://centera.tv/collateral/article/fp029-li.pdf.
[58] Kamdar, Adi, Rainey Reitman, and Seth Schoen. "NSA Turns Cookies (And More) Into Surveillance Beacons," *Deeplinks*, Electronic Frontier Foundation, December 11, 2013, accessed August 11, 2015, https://www.eff.org/deeplinks/2013/12/nsa-turns-cookies-and-more-surveillance-beacons.
[59] Englehardt, Steven, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. "Cookies That Give You Away: the Surveillance Implications of Web Tracking," *WWW 2015*, May 18-22, 2015, Florence, Italy, accessed August 11, 2015, http://www.www2015.it/documents/proceedings/proceedings/p289.pdf.

online tracking practices are deeply at odds with users' expectations.[60] [61] [62] According to Nissenbaum's theory of contextual integrity, "what people care most about is not simply restricting the flow of information but ensuring that it flows appropriately".[63] According to the Association of Internet Researchers (AoIR)'s ethical guidelines, researchers must ask "What are the ethical expectations users attach to the venue in which they are interacting, particularly around issues of privacy?"[64] Arguably, both Encore and much of third-party tracking today equally flout these expectations. In such an environment, there is a risk of an "ethical race to the bottom." Credentialed researchers and respected academic organizations should arguably not participate in and facilitate a race to the bottom even if advertisers feel obliged to do so — their tools may be similar, but their ethical obligations need not be.

Second, the probability and magnitude of harm may depend on the type of censored website. For social media sites such as Facebook, Twitter, and YouTube, which are frequently censored, an Encore measurement might not stand out in any way since widgets from these websites (such as Facebook's Like button) are encountered extremely frequently in regular web browsing. The status is less clear with other websites such as news sites, also frequent targets of censorship. The nature and magnitude of harm may also depend on the reason the website was censored. A pattern of repeated access to specific religious websites that are deemed sensitive and censored will likely be viewed differently from accesses to Facebook or Twitter. It is difficult to generalize about the feasibility of enforcing laws across different regimes with different technological capabilities, real-world enforcement resources, and, more fundamentally, different levels of respect for the rule of law. There is little information available on the likelihood and severity of persecution for simply accessing (or attempting to access) blocked domains, although of course citizens of many countries face such risks for online *writing*.[65] Tor project leader Roger Dingledine notes that there is "little reprisal against passive consumers of

[60] Ur, Blasé, Pedro G. Leon, Lorrie Faith Cranor, Richard Shay and Yang Wang. "Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising," CyLab, Carnegie Mellon University, April 2, 2012, accessed August 11, 2015, http://www.futureofprivacy.org/wp-content/uploads/Smart-Useful-Scary-Creepy.-Perceptions-of-Online-Behavioral-Advertising-.pdf.
[61] McDonald, Aleecia M., and Lorrie Faith Cranor. "Americans' Attitudes About Internet Behavioral Advertising Practices," *WPES'10*, October 4, 2010, Chicago, IL, accessed August 11, 2015, http://www.aleecia.com/authors-drafts/wpes-behav-AV.pdf.
[62] Turow, Joseph, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy. "Americans Reject Tailored Advertising and Three Activities That Enable It," *Technology | Academics | Policy (TAP),* September 2009, accessed August 11, 2015, https://www.techpolicy.com/TechnologyAcademicsPolicy/media/document-library/Americans-Reject-Tailored-Advertising---And-Three-Activities-That-Enable-It---September-2009-Survey-for-the-Rose-Foundation.pdf.
[63] Nissenbaum, Helen. *Privacy in Context Technology, Policy, and the Integrity of Social Life*, Redwood City, CA: Stanford University Press, 2009. Print. http://www.sup.org/books/title/?id=8862.
[64] Markham, Annette and Elizabeth Buchanan. "Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Committee," approved by the AOIR general membership, December 2012, accessed August 11, 2015, http://aoir.org/reports/ethics2.pdf
[65] "Freedom on the Net 2014," Freedom House, accessed August 11, 2015, https://freedomhouse.org/report/freedom-net/freedom-net-2014#.VfWZz2TBzGc.

information".[66] On the other hand, we do know that the NSA monitored visits to porn websites as part of a plan to "discredit radicalizers".[67]

Third, the focus on harm to individuals doesn't account for other types of harms that might result. For example, the authors argue that "more widespread measurements like Encore become, the less risky they are for users" by making cross-origin requests to censored domains a commonplace occurrence. On the other hand, the censors might conceivably respond by shutting down Internet connectivity altogether.

Mitigating Harm

The Encore researchers limited the set of URLs that the script induced users to measure. All such URLs came from the list that Herdict asks its users to test. The current version of Encore tests only Twitter, Facebook, and YouTube, the rationale being that these domains are accessed regularly and automatically by most users' web browsers in the course of normal web browsing. In the section on informed consent, transparency and accountability below, we discuss other (actually used as well as potential) methods to mitigate harm.

The need for harm mitigation should inform the research design process, especially in research areas where established norms don't exist. From the AoIR guidelines: "Ethical decision-making is a deliberative process, and researchers should consult as many people and resources as possible in this process, including fellow researchers, people participating in or familiar with contexts/sites being studied, research review boards, ethics guidelines, published scholarship (within one's discipline but also in other disciplines), and, where applicable, legal precedent." The document further provides several questions for investigators to consider: "How are the concepts of 'vulnerability' and 'harm' being defined and operationalized in the study? How are risks to the community/author/participant being assessed? How is vulnerability determined in contexts where this categorization may not be apparent? Would a mismatch between researcher and community/participant/author definitions of

[66] Dingledine, Roger. "Tor and circumvention: lessons learned," The Tor Project, accessed August 11, 2015, https://www.iacr.org/conferences/crypto2011/slides/Dingledine.pdf.
[67] Greenwald, Glenn, Ryan Grim, and Ryan Gallagher. "Top –Secret Document Reveals NSA Spied on Porn Habits As Part of Plan To Discredit 'Radicalizers'," *Huffington Post*, November 26, 2013, accessed August 11, 2015, http://www.huffingtonpost.com/2013/11/26/nsa-porn-muslims_n_4346128.html.

'harm' or 'vulnerability' create an ethical dilemma? If so, how would this be addressed?"

5.3 Respect for Persons, Law, and the Public Interest

Informed consent, transparency and accountability

Research ethics requires that individuals involved in a study are treated as autonomous persons. In traditional human subjects research, the investigator (ideally) approaches participants before the data collection begins, explains the research, seeks their consent. Seeking informed consent is not always feasible, as is frequently the case in network measurement research, and certain proxies for consent are sometimes deemed appropriate. Consent can be sought from a representative authority while debriefing research subjects when the data collection is completed, or the informed consent procedure may be waived completely by an IRB prior to the research. And according to the Common Rule, in cases where the researcher does not intervene in the life of an individual person to gather data, and there is no reasonable expectation of privacy, no form of consent is required. For example, anonymous observations of public activities, textual research, and examination of public records and other publicly accessible databases (even if access requires payment) are forms of research that are considered exempt from consent even though they may reveal sensitive data about persons.

The SIGCOMM program committee reviewing the Encore paper stated that the main ethical concern with the research would be mitigated if those who deployed Encore obtained informed consent from users. The authors argue against both the feasibility and desirability of obtaining informed consent and provide three related arguments. First, they say that it would be impractical since it would require teaching users "nuanced technical concepts… across language barriers" which would "dramatically reduce the scale and scope of measurements". The challenges of communicating the necessary technical information to a global set of participants again highlights the tension between the scalability imperative and established ethical norms. Second, they argue that Encore with informed consent would be essentially equivalent to existing alternatives (such as, presumably, Herdict), forfeiting the benefits of the novel measurement architecture. Third, they say that informed consent may even increase risk to users by removing plausible deniability. However, we must consider that if there is no rule of law or guarantee of a fair trial with an independent judiciary in the censoring country, plausible deniability may not help a defendant much.

Research ethics guidelines typically stress the importance of transparency of projects to serve the principles of accountability and meaningful informed consent. Additionally, guidelines recommend that "Debriefing is typically required when deception is used in order to mitigate harm resulting from loss of trust in researchers by those subjects who were deceived."

The Encore website contains a statement at the bottom “Visitors of this page have performed XXX measurements of Web filtering” and links to a page with additional information and provides the ability to opt out of future participation. Encore is meant to be deployed by other website operators; accordingly, the FAQ contains the question “Do I need to inform my site's visitors about Encore?” whose response begins “Although we cannot provide legal advice, we believe that you are not *required* to inform your site's visitors about Encore or obtain their consent before collecting measurements. That said, Encore's installation instructions explain how you can inform your visitors of Encore's presence and allow them to disable Encore entirely.” (emphasis in original). The focus appears to be on legal compliance over ethical obligation.

There are several possibilities for strengthening notice. The notice and opt-out link provided to users could be more prominent, perhaps in the style of the EU “cookie law” notices. Encore could require website operators who deploy it to give the same type of notice. Encore’s FAQ could be expanded to include an explanation of the risks and benefits of the research as well as the technical concepts necessary to fully understand these.

Legal compliance

Laws and policies regarding censorship and accessing unlawful or undesirable content on the Internet vary widely across jurisdictions; sometimes they may not be codified into law or be subject to interpretation by political officials. While the Encore team is based in the U.S., the measurement actions are triggered in browsers of Internet users worldwide, which makes the issue of jurisdiction unclear.

In terms of compliance with United States law, Encore appears to be in the clear. U.S. cybersecurity law expert Jonathan Mayer notes that “While the scope of computer abuse law remains deeply

unsettled, courts have converged on two baseline principles. First, circumventing a security protection on a remote system is illegal.[68] Second, when a system's owner explicitly revokes permission, remote access must cease.[69]" He argues that both Encore and the ZMap tool discussed in Section 2 "unambiguously abide by these guidelines" since "both measurement approaches take advantage of known, intentional software functionality." He continues, "As for respecting system owner preferences, the main deployment of both platforms is accompanied by a straightforward opt-out mechanism. If a system's owner revokes permission, research data collection immediately terminates."[70]

A global study of internet censorship law and policies — as well as other applicable bodies of law such as privacy and data protection law — would be a near impossible task for a legal researcher, let alone a team of computer scientists. Enumerating all possible (albeit remote) legal risks to Encore users is similarly infeasible. For example, the Falun Gong organization is banned in China; perhaps visits to their website may be interpreted as support for their cause.[71] Or perhaps Encore measurements are interpreted to constitute an act of espionage by helping a foreign power to map the national filter.

Since a thorough worldwide legal study is infeasible, Encore researchers cannot be certain that the measurements they induce do not constitute a violation of any local law. Ethicists would advise not putting people in a position where they could be perceived to have broken a law. In exceptional cases, however, researchers could develop an ethical justification that a law (or a type of law) is not in the public interest. The researchers must then demonstrate that they accept responsibility for their actions and the consequences, and have the necessary mitigation strategies in place [Menlo Report, p. 14].

In conclusion, Encore makes for a fascinating case study that presents a thick web of considerations and no easy answers. While the scale of today's Internet and datasets is giddying to researchers and companies alike, the ethical responsibility that comes with it is rather sobering. Our analysis reveals a complex interplay between the technical design of the experiment and its potential risks and benefits.

As of this writing, Encore is very recent work; and there is an ongoing debate about its ethics, the

[68] See, e.g., Facebook, Inc. v. Power Ventures, Inc., No. C 08-05780 JW, 2010 WL 3291750, at *5-12 (N.D. Cal. July 20, 2010).
[69] See, e.g., Craigslist Inc. v. 3Taps Inc., No. CV 12-03816 CRB, 2013 WL 4447520, at *5 (N.D. Cal. Aug. 16, 2013).
[70] Jonathan Mayer, personal communication.
[71] The full list of URLs ever measured by Encore is listed at http://encore.noise.gatech.edu/urls.html. It does not include any websites related to the Falun Gong organization.

broader question of norms for ethical research in network measurement, computer security, data science, and other disciplines, as well as the meta-question of how these disciplines should exercise ethical gatekeeping. We invite you to join the conversation.

The Data & Society Research Institute Program on Ethics in "Big Data" Research will investigate the potential benefits and challenges put forward in this primer. Through partnerships, collaboration, original research, and technology development, the program seeks cooperation across sectors to innovate and implement thoughtful, balanced, and evidence-based responses to our current and future data-centered issues.

Data & Society is a research institute in New York City that is focused on social, cultural, and ethical issues arising from data-centric technological development. To provide frameworks that can help address emergent tensions, D&S is committed to identifying issues at the intersection of technology and society, providing research that can ground public debates, and building a network of researchers and practitioners that can offer insight and direction. To advance public understanding of the issues, D&S brings together diverse constituencies, hosts events, does directed research, creates policy frameworks, and builds demonstration projects that grapple with the challenges and opportunities of a data-saturated world.

Authors:
Arvind Narayanan: arvindn@cs.princeton.edu
Bendert Zevenbergen: bendert.zevenbergen@oii.ox.ac.uk

For additional case studies:
Emily F. Keller and Jacob Metcalf
bdes@datasociety.net
Data & Society Research Institute
36 West 20th Street, 11th Floor New York, NY 10011
Tel. 646-832-2038
datasociety.net