Forgive Us our SYNs: Technical and Ethical Considerations for Measuring Internet Filtering

Jedidiah R. Crandall[1]*, Masashi Crete-Nishihata*[2]*, and Jeffrey Knockel*[1,2]

[1]*Dept. of Computer Science, University of New Mexico*
[2]*Citizen Lab, University of Toronto*

Abstract

We provide an overview of technical and ethical considerations for measuring Internet filtering, focusing on client-side and network side channel techniques. Client-side measurements are any measurement that relies on software deployed on a network vantage point. Users are typically involved in installing and running the measurement software, and informed consent process to ensure users are aware of risks in doing so should be followed. However, the relative risk is highly contextually specific, and can vary depending on the location and background of the user running the measurement. Network side channel techniques do not require installation of special software. By using measurements that take advantage of common implementation details of network stacks we can infer if Internet filtering is present between two machines. It is impractical to obtain prior informed consent for network side channel experiments. These techniques are interventions into an environment, and to be done responsibly research risk has to reduced to a minimum. We identify open questions these distinct approaches raise and discuss possible mitigations for ensuring responsible net- work measurement experiments.

1 Introduction

Network measurement research that documents Internet filtering and other forms of network interference has steadily grown in recent years [9]. As this field expands researchers increasingly find themselves in ethical grey areas and institutional gaps further complicate ethical evaluation and guidance. Research ethics standards and protocols for biomedical and behavioral research are well developed. However, ethical frameworks for research involving information communications technologies (ICTs) are not as established. For these studies the objects of analysis are often computers and networks rather than direct human subjects. The level of abstraction between a network under examination and a human user can complicate traditional research ethics principles including informed consent and evaluation of risk and benefits [11]. Efforts such as the Menlo Report have sought to address this gap and provide ethics guidance for ICT research [13]. Nonetheless grey areas and knowledge gaps persist due in part to lack of shared community values and insufficient subject matter expertise on institutional review boards [12].

Understanding the potential risk of running network measurements requires consideration of technical and contextual factors. The technical implementation of an experiment as well as the legal and political environment from which it is run can significantly alter the level of relative risk.

The majority of Internet filtering detection projects rely on client-side measurement tools and techniques[6, 8, 17]. Client-side measurements typically consist of installing measurement software on a specific vantage point. Often these vantage points are provided by research participants who are located within a county of interest.

A body of new research has shown the potential of using network side channels for detecting Internet filtering [14, 15, 19]. These techniques can be used to measure the existence of Internet filtering between two machines without the need for measurement software installed on either one. These techniques send packets to machines such as SYNs or pings with spoofed source addresses to cause those machines to automatically send responses to the spoofed addresses. By taking advantage of the common implementation details of many network stacks, these techniques can measure whether the responses were received by the machine at the spoofed address by exploiting the way that information flows throughout these machines' stacks.

Both approaches have distinct technical and ethical challenges. Client-side measurements often directly involve users and therefore the principles and process of informed consent should be followed. However, the risk of running client-side measurements is highly contingent on

the location and context of the user running the measurement, which can complicate the informed consent process. Obtaining prior informed consent for side channel measurements is not possible. These measurements are a research intervention into an environment and in order to waive consent requirements they must be done in ways that reduce risk to the minimum.

This paper provides a high level overview of ethical and technical considerations for client-side and side channel measurements. It draws on experiences from the OpenNet Initiative (ONI) [8], Information Controls Lab (ICLab) [6], and recent work on network side channel techniques [14, 15, 19]. We offer proposals for how risk can be handled and mitigated for both kinds of measurements. However, as we show, many open questions abound that require further discussion.

2 Client-side network measurement

We define client-side network measurements as any measurement that relies on software deployed on a network vantage point. A number of projects focused on detecting Internet filtering [6, 8, 17] and network interference and more general network performance measurements [7, 20, 21] rely on this approach.

In the following sections we provide an overview of the techniques, trade-offs, and research risks that client side measurements can pose for network interference studies. We conclude with a proposal for how to develop contextually aware informed consent that can provide better guidance to users. These sections draw from the experiences of the ONI and ICLab.

2.1 OpenNet Initiative client-side measurements

The OpenNet Initiative, was an inter-university consortium, that measured national-level Internet filtering through an interdisciplinary approach that combined network measurements with fieldwork, legal, and policy analysis. The project ran from 2003-2014 and conducted measurements in 77 countries and found that 42 of them – including both authoritarian and democratic regimes – implement some form of filtering [18].

ONI utilized a client-based in-country testing method for network measurement. This approach uses software written in Python in a client-server model, which is distributed to researchers. The client attempts HTTP GET requests to a pre-defined list of URLs simultaneously in the country of interest (the "field") and in a control network (the "lab"). After measurements complete, results are compressed, and transferred to a server for analysis. Data collected by these measurements include HTTP headers and status code, IP address, page body, and in some cases traceroutes and packet captures [18].

In this approach, implementing a measurement consists of the following steps: 1) the country and network of interest are identified; 2) a URL testing sample is developed usually in collaboration with researchers from the country of interest or who otherwise have specific area studies expertise. Typically two lists are used: one with globally sensitive URLs tested in every country and one with locally sensitive URLs selected for the specific country; 3) a user is identified. In the ONI these users were often collaborators in the consortium who used the resulting data for research and advocacy activities; 4) the user goes through an informed consent meeting where the research risks and objectives of the project are explained. This meeting also includes probing the user about their perception of the relative risks of running measurements in the target country. As the users were often more familiar with the context of the specific country their perceptions of risks were an important input into determining if and how an experiment could be run safely; 5) if the user consents to the study they would install and execute the software client according to a specific testing schedule.

2.2 Client-side measurement risks

The HTTP GET requests the client initiates are done without any obfuscation or anonymization as the intent is to collect responses from the network that are representative of the average user experience. This method introduces risk. If the users' network is being monitored requests to potentially sensitive URLs can be discovered.

While technically the risk can be explained the exact level of it is highly contextually specific. Depending on the location of the vantage point there could be restrictive legal and regulatory frameworks and government authorities who could potentially see this research activity as representing a challenge to state authority, as the intention is to report on the existence, depth, and prevalence of national filtering systems.

Social risk (e.g. loss of status, privacy and / or reputation) could potentially emerge if external agents become aware of a research participant engaging in the study. For example in some countries authorities may view this form of research as subversive. If an authority is sensitive to information of its censorship regime being exposed and views exposure as potentially damaging, nationals of that country who are found to be engaged in this study may be scrutinized.

Assessing the legal risks involved is also contextually specific. In some countries (e.g. Iran) it is illegal to circumvent Internet filtering [1]. However, to our knowledge there are no explicit laws that forbid the collection of measurements on network interference. Despite the lack of explicit laws that prohibit network measurements and the fact that our measurements do not harm or disrupt the networks under investigation the nature of this research can potentially have political impacts and therefore could

be seen unfavorably by certain authorities. In countries with problematic rule of law and records of targeting dissent it is possible that nationals engaging in research with a foreign university to collect data on a country's Internet filtering regime could be targeted by authorities and potentially have legal action taken against them (although it is unclear to us what exact form this legal action could take).

Concerns surrounding safety and practicality drove decisions of where the ONI did measurements. Often the ONI considered countries with potential for interesting data as too risky or impractical for user-based testing (for example, during the recent conflict in Syria, and in countries like Cuba and North Korea). Identifying the high threshold for user risk can be straightforward, but providing accurate user guidance in every situation is challenging and requires focused interactions between the user, the researchers, and others with area expertise.

In over 10 years the ONI never experienced a user being arrested, apprehended, pressured, or intimidated by authorities for their participation in the project. However, this lack of experienced negative impacts does not mean that risk of legal or extralegal persecution does not exist.

2.3 Client-side measurement tradeoffs

Client side measurements have limitations. In many countries, vantage points capable of running measurement software are either unavailable or insufficiently geographically diverse. In other countries the research risk for users may be too high to run measurements on user controlled machines located within a country's jurisdiction. Thus, the scalability of client side measurements can be restricted by availability of vantage points and safety and ethics concerns.

2.4 Responsible client-side measurements

Client-side measurements involving users should go through the informed consent process and ensure user participation is both free and informed. As we have discussed the challenge is the potential variability in risk that is contingent on the specific location of the measurement.

As part of the Information Controls Lab (ICLab) we are developing means to facilitate contextually aware informed consent. At a high level the process for running measurements in ICLab is similar to ONI. The project protocol has been reviewed and approved by the Research Ethics Board at the University of Toronto. As part of the project we are exploring how to assess varying levels of risk and communicate them to users during the informed consent process.

We assess the research risk a country poses using the following metrics: Freedom House "Freedom on the Net score" (a study of Internet freedom in 47 countries that are based on laws and practices relevant to Internet accessibility and rights) [4]; Economist Democracy Index (analysis of economic, financial political and business risk of 203 countries that categorizes the countries' government into four types: Full Democracy, Flawed Democracy, Hybrid Democracy and Authoritarian) [3]; Government of Canada travel advisory (alerts on situations that can affect wellbeing and safety abroad) [5].

These baseline scores are used to determine if a country falls within low, medium or high risk. We are restricting our study to countries that present the medium research risk level. Potential research participants who request to engage in the study but are from countries that present high risks (e.g. active conflict zones) are considered ineligible to participate.

For example countries such as the US which score "Free" on the Freedom on the Net metric, "Full Democracy" on the Economist Democracy Index and have no travel warnings or other reports of caution would be assessed as "low research risk". A country like Pakistan which scores "Not Free" on the Freedom on the Net metric, "Hybrid democracy" on the Economist Democracy Index, and a warning to avoid nonessential travel on the travel advisory would be assessed as "medium research risk". A country like Syria that scores "Not Free" on the Freedom on the Net metric, "Authoritarian" on the Economist Democracy Index, and has a travel warning to avoid all travel would be assessed as "high research risk". These baseline scores will be further informed by reports from the ground that will be regularly provided by our research team and network. We will have the option of manually adjusting a risk score if a sudden development occurs in a country that would not be covered by our metrics (e.g. mass protests, conflict, etc).

We will also consider how the selection of URL samples affects the relative risk in a particular jurisdiction. For example, in some environments testing for content related to child abuse and terrorism may be further scrutinized by authorities as accessing this content is explicitly prohibited.

We are currently piloting these metrics with users from countries around the world at varying risk levels (excluding high risk). We acknowledge that these metrics are not perfect approximations of risk, but we hope that can serve as a baseline and be informed by further context.

Ideally, we envision a database that is regularly updated with standardized metrics, vetted situation reports from recognized experts and user updates that provide a more holistic view of potential risks in countries under consideration for network measurements.

3 Side-channel network measurement

Side channel techniques can be used to measure censorship between two machines without running special measurement software on either one. These techniques

send packets to machines such as SYNs or pings with spoofed source addresses to cause those machines to automatically send responses to the spoofed addresses. By taking advantage of the common implementation details of many network stacks, we can measure whether the responses were received by the machine at the spoofed address by exploiting the way that information flows throughout these machines' stacks.

In the following sections we provide an overview of the techniques, technical trade-offs, research risks, and open questions regarding how to run responsibly run these measurements.

3.1 Side-channel network measurement techniques

An idle scan is a side channel technique that takes advantage of the fact that some machines' network stacks fill the IP Identifier (IPID) field in the IP header using a globally incrementing counter [15]. This counter enables us to, for example, send two pings to such a machine and then count how many packets the machine has sent between the two pings by taking the difference of the IPIDs in our ping responses. By sending a SYN to a server with the source address of a desired vantage point with a globally incrementing counter, we can measure whether the server's SYNACK response was received by the vantage point by measuring whether the vantage point sent a RST to the server in response to the SYNACK. One disadvantage of an idle scan is that vantage points, in addition to having globally incrementing IP id counters, must be idle; otherwise, it is difficult to determine whether the client is sending RST's or if it is sending packets to other hosts.

Another technique called a backlog scan, takes advantage of the fact that servers' SYN backlogs have finite storage space [22]. We first send the server a large number of "canary" SYNs to fill a large amount of the backlog. Then we send the server SYNs with the spoofed source addresses of our desired vantage point as with the idle scan. If neither of the SYNs' corresponding SYNACK or RST packets were filtered, then these SYNs will be removed from the backlog when the server receives the RST. Otherwise, the SYNs will remain in the backlog, taking up space. Finally, we send the server another large number of canary SYNs. Assuming we sent enough canaries to overflow the backlog, then some of our canaries that we initially sent will have been evicted; however, if the spoofed SYNs were never removed from the backlog, then even more of our canaries will have been evicted. We can count how many of our SYNs are still in the backlog by attempting to complete the handshake all of these half-open connections; however, other packets can be sent to count them without having to complete the handshake as described in Zhang et al [22]. By comparing the number of SYNs still in the backlog, we can measure whether the spoofed SYNs were removed from the backlog or if there was filtering.

A third technique is called a fragment scan. This technique is similar to the backlog scan except that it instead takes advantage of the fact that machines have finite storage space for their fragment caches that they use to collect incoming fragmented IP datagrams' fragments. Unlike the previous scans, this scan operates in Layer 3 and thus can be combined with one of the previous Layer 4 techniques to distinguish between censorship implemented in Layer 3 versus Layer 4.

3.2 Side-channel research risks

Unlike client-side measurements it is not possible to receive prior informed consent for side channel experiments. Research ethics frameworks state that the following conditions must be met for informed consent to be waived: 1) the research involves no more than minimal risk to participants; 2) the waiver of consent requirements is unlikely to adversely affect the welfare of the participants; 3) it is impossible or impractical to carry out the research and address the research questions properly if prior consent of participants is required [13].

If the intent of a project is to use one of the side channel techniques we outline above to measure Internet filtering then obtaining prior informed consent is impractical, as it would require having some means to communicate with the users beforehand and would reduce and bias the random sampling that is desirable to get a wide range of representative measurements. Given the impracticality of obtaining prior informed consent side channel measurements must be designed with technical mitigations for reducing risks to a minimum and protect the welfare of any affected stakeholders.

As many of these techniques are novel and have not been extensively studied (especially in the context of Internet filtering detection) there are numerous open ethical and technical questions. In the following section we outline technical and ethical tradeoffs these techniques present and the open questions that emerge.

3.3 Responsible side-channel measurements

Table 1 lists common side channel techniques and their respective tradeoffs, with respect to both measurement concerns and ethical concerns. For illustration purposes, we assume that any given side channel measurement is performed from a measurement machine under the control of the researchers, and the goal of the measurement is to detect filtering between a given client and server that can be anywhere on the Internet. For side channels not based on TCP/IP the notion of server and client is not important, they can just be any two machines, but we use the terms client and server for all side channels in this

Side Channel	Measurement Considerations	Ethical Considerations
IPID	Approx. 1% of IPv4 address space has global IPIDs; difficult to apply in IPv6 because the fragment ID is only included in fragments	Requires a relatively low rate of packets (e.g., 5 per second); measurement machine must communicate directly with client.
SYN backlog	Every machine on the Internet with an open port (i.e., every server) has a SYN backlog; SYN backlogs vary from OS to OS; relatively noise-free signal compared to IPID side channel	Some OSes do not properly protect themselves against DoS; requires a relatively low rate of packets (e.g., 5 per second); measurement machine must communicate directly with server.
Fragment cache	Virtually no noise; fragment cache implementations vary widely	Some fragment cache implementations do not protect properly against DoS; packet rates depend on fragment cache size.
ICMP rate limits	Not well studied; may be relatively low noise.	Can be elicited from gateway routers for IP addresses where no machine exists; may be possible to use at very low packet rates (e.g., 5 per second).

Table 1: Technical and Ethical considerations for Side Channel Techniques

discussion.

The IP Identifier (IPID) side channel, known as an idle scan, was first proposed by Antirez [10]. It is relatively well understood in terms of its noise properties, and is widely applicable since about 1% of the address space has global IPIDs and even for machine without global IPIDs some information flow exists in the IPID.

The SYN backlog idle scan was first presented in Ensafi et al [15] in a form that required denial of service (DoS) for the side channel to be used, but more recent implementations only require that the SYN backlog be half full for information to flow [16, 22].

The hybrid idle scan is not listed in Table 1, but it can be seen as a combination of the IPID side channel and the SYN backlog side channel. This technique can additionally allow the direction of filtering to be measured with the same tradeoffs of using the IPID side channel.

The fragment cache side channel was first presented by Knockel and Crandall [19]. It is the most widely applicable side channel in terms of the high number of Internet hosts that will reply to IP fragments, including routers and machines that are otherwise completely inaccessible because of a host firewall. However, the implementations of fragment caches for various OSes and devices varies widely.

ICMP rate limitations were first presented by Ensafi et al [15]. There has not been enough research into how hosts on the Internet respond with ICMP, what rate limits apply, and how those rate limits are applied to fully understand this side channel.

3.4 Open ethical questions

We now discuss side channel measurements techniques in the context of the following ethical questions:

Is it necessary for the measurement machine IP address to appear in traffic logs for both the server and the client? It is common practice for Internet measurements to host a web page on the measurement machine stating the purpose of the measurements and providing contact information for network administrators that would like to "opt out" of future measurements. This practice assumes that the measurement machine's IP address appears in packet logs for all machines and networks that are targeted in the measurements. In order to perform side channel measurements, the measurement machine generally must communicate directly with either the client or the server. However, of the side channels discussed above only the hybrid idle scan and the fragment cache side channel have the property that the measurement machine communicates with both the server and the client during measurements. For the other side channels, the measurement machine could be made to send packets to both the client and server but this may make it much easier for censors to block measurements.

What is the proper level of risk with respect to denial of service? In general, all of the above side channels can be utilized without causing denial of service, assuming that only one measurement machine is performing measurements at a time for any given server or client. Two exceptions are that some machines (e.g. some versions of Windows) do not properly protect themselves against SYN floods, meaning the SYN backlog side channel causes DoS in these cases, and that some small fraction of fragment cache implementations do not properly protect themselves against DoS when their fragment cache is filled by measurements using the fragment cache side channel. In most cases, even if DoS occurs it is easy to detect and relatively minor. For example, virtually all modern network stacks will send SYN cookies by default when the SYN backlog becomes full. SYN cookies still allow other clients to connect, but with sometimes reduced throughput (because of the lack of a scaled flow control window). Furthermore, it is possible for measurement machines to detect other measurements being

performed and back off. All of the above side channels can be developed and adapted to mitigate DoS, but what level of risk with respect to DoS should be designed into side channel measurements?

What packet rates are acceptable? A related, but separate, question is: what packet rates are acceptable? This number can also depend on the type of packet. For example, SYN packets at a high rate that does not cause DoS can still be flagged as a potential attack by a network intrusion detection system. Large ICMP packets sent at a relatively slow rate can use a lot of bandwidth and potentially consume the bandwidth on low-bandwidth links.

What kind of machine should the client be? What if the client is associated with a person? A major issue that arises from side channel measurements is that they may intervene in the client's environment and cause risk for the client. A specific concern is that an individual or government monitoring the network will mistakenly think that the client is trying to communicate with the server. An idea (that has been mentioned by a number of colleagues) is to traceroute to a given client and then find a router near the client that can be used as the client instead, e.g., a router with a global IPID can be used as a client in the hybrid idle scan. This mitigation is what is done in the Censored Planet project [2]. Another possibility is to use something like the ICMP rate limitation side channel, where the "client" can be an IP address that is unresponsive (i.e., no client exists at that IP address) and the side channel is actually being measured on its gateway router. Yet another possibility is to use servers as clients, on the theory that web servers cannot be tied to one person as clients. For noisy side channels such as the IPID or hybrid idle scan, this may not be practical because servers typically have a high amount of IPID noise.

Another possible mitigation is to perform measurements for entire /24s at once, so that no individual can be associated with the measurements incorrectly. Any side channel measurement can be performed in this way, but this technique will be more effective if a large fraction of the /24 meets the requirements of the measurement. For example, the fragment cache side channel could be performed usefully for a large fraction of many /24s.

4 Conclusion

Client-side and side channel measurements should be seen as options in the network measurement toolkit. Deciding which approach is best for a particular research question requires careful assessment of technical and ethical considerations. As open questions regarding responsible measurements abound the community should see research ethics as both a practical concern and a subject of study on its own. Guidelines and protocols for running experiments need to be developed from within the community and inform institutions on how accepted research ethics principles apply to our work and what grey areas still persist.

Acknowledgments

Jeffrey Knockel was supported by the Open Technology Fund Information Controls Fellowship Program. The research of the Citizen Lab was supported by Social Sciences and Humanities Research Council of Canada (SSHRC) grant 430-2014-00183. This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. #1314297 and #1420716. We would like to thank Nicholas Weaver for valuable insights about the ethics of the ICMP rate limitation side channel.

References

[1] Available at http://www.cyberpolice.ir/page/2551.

[2] Censored Planet. Available at http://www.cs.princeton.edu/~rensafi/projects/Censored-Planet/index.html.

[3] Economist Intelligence Unit, "Democracy Index". Available at http://www.eiu.com/democracy2014.

[4] Freedom House, "Freedom in the World 2015". Available at http://https://freedomhouse.org/report/freedom-world/freedom-world-2015.

[5] Government of Canada, "Country travel advice and advisories". Available at http://travel.gc.ca/travelling/advisories.

[6] Information Control Lab. Available at https://iclab.org/.

[7] Measurement Lab. Available at http://measurementlab.net.

[8] OpenNet Initiative. Available at https://opennet.net.

[9] Aceto, G., and Pescape, A. Internet censorship detection: A survey. *Computer Networks* (2015).

[10] Antirez. new tcp scan method. Available at http://seclists.org/bugtraq/1998/Dec/79, 1998.

[11] Deibert, R., and Crete-Nishihata, M. Blurred boundaries: Probing the ethics of cyberspace research. *Review of Policy Research 28*, 5 (2011), 531–537.

[12] Dittrich, D., Bailey, M., and Dietrich, S. Building an active computer security ethics community. *IEEE Security and Privacy 9*, 4 (July 2011), 32–40.

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

[14] Ensafi, R., Knockel, J., Alexander, G., and Crandall, J. R. Detecting intentional packet drops on the Internet via TCP/IP side channels. In *proceedings of the 2014 Conference on Passive and Active Measurements* (Albuquerque, New Mexico, 2014).

[15] Ensafi, R., Park, J. C., Kapur, D., and Crandall, J. R. Idle port scanning and non-interference analysis of network protocol stacks using model checking. In *Proceedings of the 19th USENIX Security Symposium* (2010), USENIX Security'10, USENIX Association.

[16] Ensafi, R., Winter, P., Mueen, A., and Crandall, J. R. Analyzing the Great Firewall of China Over Space and Time. In *2015 Privacy Enhancing Technologies Symposium* (2015), Springer.

[17] Filasto, A. OONI: Open observatory of network interference. In *2nd USENIX Workshop on Free and Open Communications on the Internet* (Bellevue, WA, 2012), USENIX Association.

[18] Gill, P., Crete-Nishihata, M., Dalek, J., Goldberg, S., Senft, A., and Wiseman, G. Characterizing web censorship worldwide: Another look at the Opennet Initiative data. *ACM Trans. Web 9*, 1 (Jan. 2015), 4:1–4:29.

[19] Knockel, J., and Crandall, J. R. Counting packets sent between arbitrary internet hosts. In *4th USENIX Workshop on Free and Open Communications on the Internet (FOCI 14)* (San Diego, CA, Aug. 2014), USENIX Association.

[20] Kreibich, C., Weaver, N., Nechaev, B., and Paxson, V. Netalyzr: Illuminating the edge network. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (New York, NY, USA, 2010), IMC '10, ACM, pp. 246–259.

[21] Sánchez, M. A., Otto, J. S., Bischof, Z. S., Choffnes, D. R., Bustamante, F. E., Krishnamurthy, B., and Willinger, W. Dasu: Pushing experiments to the internet's edge. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)* (Lombard, IL, 2013), USENIX, pp. 487–499.

[22] Zhang, X., Knockel, J., and Crandall, J. R. Original SYN: Finding machines hidden behind firewalls. In *Proceedings of INFOCOMM 2015* (2015).