Submission Summary ACM SigComm2015 - Workshop on Ethics in Networked Systems Research

Considerations from practical examples

Chokepoint Project
info@chokepointproject.net

Wednesday 1st April, 2015

1 About The Chokepoint Project

The Chokepoint Project's mission is to perform continuous monitoring of network behaviour in conjunction with legal data, journalistic events and econometric data to provide up to date and contexualized information about human and civil rights in a digital environment.

The project processes third party open data as well as harvesting data from network probes run by the project. For more information about The Chokepoint Project, visit: `https://beta.chokepointproject.net/about/`

2 The Mysterious Case of the Disappearing Domain

Early February 2015 we received signals about problems with access to the website of Sri Lankan newspaper Colombo Telegraph, `www.colombotelegraph.com`. This prompted us to investigate more closely. [1]

The Chokepoint Project's DNS monitoring system [2] is based on DNShonest, a research tool developed by Dr. Joss Wright of the Oxford Internet Institute. The monitoring system runs every 12 hours, querying a selected set of DNS

[1] https://chokepointproject.net/the-disappearance-of-flight-www-colombotelegraph-com/
[2] https://beta.chokepointproject.net/measurements/suspicious-dns-responses

servers for a selection of domain names, and classifies the responses based on how likely they are to represent interference. [3]

The results from the monitoring system provided strong indication that the DNS was being tampered with to prevent people from reaching the website[4]. Shortly after verifying that there was clear evidence of intentional blocking, on February 5, 11:42 UTC, the website became accessible again. In the weeks to follow we also saw a reemergence of the other previously DNS filtered sites.

3 All is well in the state of Denmark?

It is an established fact that DNS manipulation has been used to implement censorship policies[3][1][2]. Limiting access to information by manipulating the DNS is easy and cheap, since it relies on network infrastructure that is already in place and requires no additional knowledge beyond what is needed to set up and manage the network in the first place. As such, it is not surprising that this method was used to suppress, perhaps, some inconvenient fact or exert some pressure. For us, it is gratifying to think that our research may have contributed to lifting the blockade a bit sooner than otherwise might have been the case, and for the Colombo Telegraph, the outage, while intimidating, also generated significant media attention.

All is well in the state of Denmark. But what's that smell?

4 Much networks, many ethics. So Denmark.

Our involvement in the Colombo Telegraph incident highlighted a number of ethical concerns:

- A single run of the DNS monitoring system for Sri Lanka issues about 180,000 DNS requests. This scans only a small section of the available domain names and DNS servers. A reasonably comprehensive scan would require scaling this up by several orders of magnitude. Is it ethical to use network resources and third party equipment in this way?
- Aside from the Colombo Telegraph domain, our monitoring system identified intervention with several other domains, including porn sites and one site with racist and antisemitic 'hate speech'. In our press release, we did specifically mention the domains hosting 'adult content', but did not mention the domain propagating hate speech. How do expectations of transparency – broadly viewed as a *sine qua non* for any form of democratic participation – affect our interest to not have our platform advertise views we find personally repugnant?

  Mirroring this, we also identified one ISP from which no interference was detected. While we could not reliably determine if this was due to nameserver IP selection not being reliably representative, we might simply not have been monitoring nameservers from that ISP that were frustrating

[3] https://beta.chokepointproject.net/static/publications/dnshonest_cpp.pdf

[4] Specifically, access was denied by having DNS servers return the localhost address, 127.0.0.1, for this domain name, as well as others.

access. Before realizing this possibility we did consider explicitly mentioning this ISP as a form of 'positive reinforcement', but elected not to do so because our 'compliment' might have drawn explicit attention to what might have been an expression of civil disobedience.

- The addresses of the Sri Lankan domain name servers that were used to perform monitoring runs were acquired based on the Internet Census data. The Internet Census data can be criticized for the problematic viral method of probe distribution. But openly resolving DNS server, the extract used by The Chokepoint Project, are publicly discoverable and could have easily been required without resorting to viral probe distribution. Irrespective of the probing method, or even of that which is probed, there are a number of fundamental questions:

  1) Where do the rights of network owners and administrators to control their infrastructure impede on effective controls to safeguard the rights to privacy and freedom of expression?

  2) Does the inherent openness of network behavior ipso facto mean that continuous collection of such behavior also make ethical and moral acceptability inherently applicable? I.e. does non-invasiveness of a measurement methodology, or inherent openness of a data element, have any relevant impact on the qualification of such a monitoring activity as surveillance, mass or otherwise?

5 Discussion

When we started to work on The Chokepoint Project we formulated a number of general principles with regards to data collection and harm reduction, such:

- Be verifiably true. I.e. one source is no source.
- Do no harm. I.e. does not contain information which could conceivably result in harm to an individual or group.
- Be verifiable. I.e. used data should be publishable or publicly accessible.

This led to 'rules', such as:

- Raw data, usage or which could endanger others, is not be used.
- Anonymization of data has to be done at the source. I.e. not by us.
- Data from a single source is not to be used.
- We will not directly engage individuals to perform tests for us.

Over time we have had to reinterpret these 'rules' as 'rules of thumb', as they proved impractical in their strictest interpretation. What defines a 'single' source in the context of data acquisition can already be problematic. Take Mlab as an example, despite the origin of that data being the participants that run the tests, from our vantage point mlab is our 'single' source for much of the data types they provide. Additionally not all mlab's data is anonymized at the source, as such is not a direct requirement for the intent of the tests. This leaves us with three choices 1) to not use mlab data, 2) to anonymize already public data or 3) to ignore the issue altogether. None of these are optimal, clearly.

To 'do no harm', suggests an understanding of what harm might be inflicted. While many scenarios can easily be imagined, we are continuously troubled by what we know not to know: every cultural, political and legal sensitivity of every country or region on the planet.

Networked systems research, when not exclusively in the realms of property rights, other economic considerations or engineering, seems almost unavoidably afflicted with a political or 'activist' dimension. Already before the publication of an analysis, automated or otherwise, a choice of what to monitor or measure is intrinsically linked to a researcher's political, cultural, even moral, framework.

References

[1] Anonymous (2014) *Towards a Comprehensive Picture of the Great Firewall's DNS Censorship*, 4th USENIX Workshop on Free and Open Communications on the Internet (FOCI 14)

[2] Dornseif, M. (2004) *Government mandated blocking of foreign Web content*, Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Editors) Security, E-Learning, E-Services: Proceedings of the 17. DFN-Arbeitstagung ueber Kommunikationsnetze, Duesseldorf 2003, ISBN 3-88579-373-3; Series: Lecture Notes in Informatics ISSN 1617-5468; Pages 617-648

[3] Wright, J. (2014) *Regional Variation in Chinese Internet Filter ing. Information*, Communication & Society 17 (1) 121-141.