How can you tell what's banned on the internet?

Joss Wright describes the technical and ethical challenges in investigating online censorship.

The belief in the internet as an unregulated stronghold of free expression and access to information remains surprisingly persistent. While we are aware of the internet's potential for pervasive mass surveillance, and have an increasing awareness of companies such as Google and Facebook turning our private lives into profit, we still happily transmit some of our most intimate details and moments online.

Globally, the role of the internet in restricting rather than providing access to information is well known. The filters employed by countries such as China, Iran, and Turkey are common knowledge. What is reported much less widely is that in many countries around the world, including the UK and others in Europe, the means to restrict access are being developed and

actively employed.

Activists, hackers, and academics have been studying online censorship for more than a decade. The activist and technical communities have largely focused on measuring filtering and developing circumvention technologies: What content and keywords are being blocked in Iran? How can internet users in Turkey access YouTube when the latest ban comes into effect? How can EE mobile users in the UK bypass overly-restrictive adult filters?

More recently, and to a lesser extent, broader questions are being asked about censorship: Why is a particular topic filtered whilst others are ignored? How do circumstances, such as political unrest, alter the topics and severity of blocking? As an active choice on the part of both national and private censors, blocking provides insight into the motivations of those who control the network. At the Oxford Internet Institute I am leading a research project, currently funded by Google, in which we aim to answer these questions using tools from both computer science and the social sciences.

A major concern in such research is to gain access to sufficient reliable data concerning censorship practices. In some senses, researching internet filtering is simple: try to access the Pirate Bay from a UK-based internet connection and you are likely to be unsuccessful—a 2012 High Court ruling required the six largest ISPs in the country to block the site, and such blocking is easy to detect. What's much harder, from an internet connection in the UK, is seeing what is blocked for a user in Shenzhen, Almaty, or Sana'a.

Traditional means to investigate media freedoms can give the greatest insights into censorship. Internet measurements, no matter how sophisticated, cannot rival the contextual knowledge of a local expert. It may be possible to detect that Facebook is blocked, but how that blocking is portrayed, how it is rationalised, and how blocking fits into culture and politics are questions that require human answers.

Human research, however, carries both costs and risks. While local experts can be a significant advantage, the need to build a relationship with appropriate partners limits this approach to focused, targeted cases. To investigate responses to emerging events, such as Egypt's almost total severing of its international connections during the revolution in 2011, establishing a local network of technical experts in only a few days is not a reliable strategy.

A further complication is that detection of censorship usually relies on trying to access banned content: a block on Facebook is detected by trying to access Facebook. For these cases, such attempts may be unlikely to lead to harm, but investigating blocking of discussion forums for homosexuals in countries where such acts are illegal, or testing access to extremist forums in the UK, may carry severe penalties. Especially where risks are unknown, or unpredictable, there are serious ethical constraints when conducting network experiments via human proxies.

Some approaches to researching censorship use software that automatically tests for filtering, and

make that software freely available for users to download. With enough users, this can provide detailed and thorough coverage, but carries even more significant ethical concerns. A user may be happy to help investigate online freedoms by installing software, but be much less happy to learn that the software is automatically connecting to blocked or illegal websites on a regular basis.

Censorship can be investigated more directly through proxy services such as Virtual Private Networks, or the anonymising Tor network. Understandably, but unfortunately, many such services aim to bypass rather than study censorship; the result being that many public proxies provide a view of the internet from the perspective of the US or Europe, but very few allow the same for, say, North Korea.

A more technically sophisticated approach is to take advantage of existing internet services in order to glean information about network manipulation. Experimenting with some common services, such as the Domain Name Service (DNS), can reveal a great deal about how and where censorship takes place. Similar techniques provide information by exploiting peculiarities in how core internet protocols work. At the Oxford Internet Institute, our research with these approaches has allowed us to study variations in censorship from place to place in China, and to discover trends and patterns in the network's filtering behaviour over time.

We still don't know exactly why some controversial traffic in China is entirely cut off while other, seemingly identical, traffic is re-routed to computers in Beijing. We still aren't entirely sure why a proportion of the traffic destined for the Tor Project's website is instead redirected to a pet grooming service in Florida. What we believe strongly, however, is that the internet's potential as a tool for control rivals its ability to provide access, and that we must understand and challenge this trend if the internet is to continue to promote, rather than hinder, free expression.

Dr. Joss Wright is a research fellow at the Oxford Internet Institute, University of Oxford. His work focuses on the investigation of online censorship and the development of privacy enhancing technologies.

Published on:October 21, 2014