MAKING TRAFFIC PUBLIC

A PROPOSAL FOR A PUBLIC STUDY OF INTERNET USAGE IN CANADA
BY FENWICK MCKELVEY

Introduction

Canadians, for too long, have been excluded from important decisions about the future of the internet, in part, because vital data about its usage remains hidden from the public eye. How much capacity do our networks hold? How much congestion exists on them? How great is the divide between urban and rural broadband? The lack of a clear picture of the internet thwarts public participation in the debate concerning its public good. This chapter imagines a public research project to collect and share information about how we use the internet.

Canadians have a rich history of broadband advocacy[1] and recent developments suggest that the means exist to mobilize this enthusiasm into a public research project. Millions of Canadians already create data on the web each day. Our posts to Facebook, edits of Wikipedia, and stream of tweets contribute to the content driving the most popular sites online. Interactive forms of research and data collection now offer a solution to the lack of public data about internet usage.[2] With the assistance of new online tools, Canadians could monitor their own internet usage, pool this data in a public resource, and analyze it to better understand how the internet operates in Canada. The objective is to facilitate greater public participation in the political and policy processes through the production and analysis of internet usage data.

What are internet usage data?

Internet usage data helps us understand how we use the internet. Usage differs from related questions of internet access. How many Canadians have access to the internet? How much does access cost and is it affordable? Does internet service provision in Canada exclude, for example, First Nations, the elderly, the poor, or rural and remote communities? The answers reveal many of the digital divides that exist in our society. Internet usage, on the other hand, asks what we do *with* access to the internet. Answers reveal another set of digital divides – technical barriers, bandwidth issues, traffic shaping, and access filtering. Internet usage also explores the popular activities online. What are the most popular sites or ways of communicating online? Since the internet is still developing as a medium, usage describes how we have come to use the web and guide its future direction.

Many indicators exist throughout the world to explain internet usage but different methods are used to produce them. The Organisation for Economic Co-operation and Development (OECD), for example, surveys its member governments to compare their national average advertised download speeds, and the types of usage limits on monthly plans.[3] Other studies provide more specific information about internet traffic. The Canadian Internet Use Survey 2009, which surveyed 23,000 Canadians, showed that email remains the most popular application among this group.[4]

Many methods of measuring usage exist because making the internet understandable remains a challenge due to its complexity. Translating the complex mess of wires, machines, humans, and

software into trends about bandwidth remains difficult and often prone to framing – i.e "select[ing] some aspects of a perceived reality and mak[ing] them more salient in a communicating text".[5] What aspects of the internet do certain usage data emphasize?[6] Often who studies the internet engenders particular ways of understanding it. The challenge, in short, involves finding accountable and accurate representations of communication online.

How does this relate to policy in Canada?

Despite the difficulty of measuring and presenting internet usage, it exerts considerable influence in our communications policy-making process. The CRTC recently ruled on acceptable internet traffic management practices (ITMPs) for Internet Service Providers[7], and usage data played a major role in the proceedings. Participating parties seeking to influence the ruling constantly cited conflicting internet usage data. The public had limited access to the data cited – keeping the debate outside of the public view. Bell Internet, for instance, provided congestion data for their networks to make the case for the growing threat of network overload, but they filed this data in confidence with the CRTC.[8] To be fair, releasing data about their networks might threaten their competitiveness. Yet, this secrecy means the public lacks information to counter the claims and hold these firms accountable.

When alternative data on internet usage do appear, they often counter the accuracy of the data submitted by the incumbent telecommunications firms. Incumbents in the CRTC hearings complained about the overwhelming growth in BitTorrent traffic, a popular peer-to-peer file sharing application. BitTorrent countered the claims of Rogers Communications spokesperson Ken Engelhart who argued peer-to-peer traffic caused congestion as it "takes place 24 hours a day seven days a week at the maximum rate of speed that the customer's service permits".[9] BitTorrent data, collected when a client "starts up or has been on/active for 24 hours", found "the average client is 'on' or active for 10-20% of the days of any given month".[10] Hardly the constant usage cited by Engelhart.

The examples above illustrate how internet usage data plays a major role in policy formulation, but the production and distribution of this data hamper public involvement in the policy process. The conflicting and confidential data on traffic management left the CRTC to sort out the mess away from the public eye. Better public data on this issue might have aided the CRTC decision – at least making it more public. It certainly would contribute to public participation in any future policy development.

Past public projects on internet usage

The transparency issues and lack of public data could be resolved by more public research on internet usage. We already have some examples of such projects.

ISPs in the United States and Canada only reluctantly admitted to traffic shaping practices after concerned media reform activists made these practices public. In 2007, the Electronic Frontier Foundation (EFF) and the Associated Press (AP) monitored BitTorrent traffic on the network of American ISP ComCast and detected it deliberately injecting 'reset' packets into this traffic.[11] Deep packet injection, as they called it, disrupted BitTorrent communication by causing the computer on one end to think the machine on the other end had hung up. The practice allowed ComCast to diminish BitTorrent traffic on their network. EFF discovered the traffic shaping using a free software packet inspection tool. Their findings prompted an investigation by the United States Federal Communication Commission.[12]

Where the EFF and AP study focused on one ISP, the *Vuze* BitTorrent application sought to understand the impact of traffic shaping on internet usage by asking its users to install a plug-in to monitor their traffic and send the results to Vuze for analysis. Eight thousands users responded and logged 100,000 hours of traffic usage data.[13] With this data, Vuze created a list of the 'Bad ISPs' that throttled traffic.[14] Many of the ISPs on the list had not widely publicized their traffic shaping, especially in Canada. The list ranked Canada's Cogeco as the second worst offender. The revelation spread through the news, provoking public concern that fueled the CRTC's hearings on ITMPs.[15]

The work of EFF, AP, and Vuze illustrate how a public research project could operate. Both studies depended on 'crowdsourcing' as an alternative to user or government surveys. Crowdsourcing, a word popular in business literature, refers to "ways to tap the latent talent of the crowd".[16] This tactic is also used in the Herdict Web project which studies internet censorship by asking its users to report the times they cannot access a website and their location. The site creates a map of the world with updates appearing as users report blockages. [17]
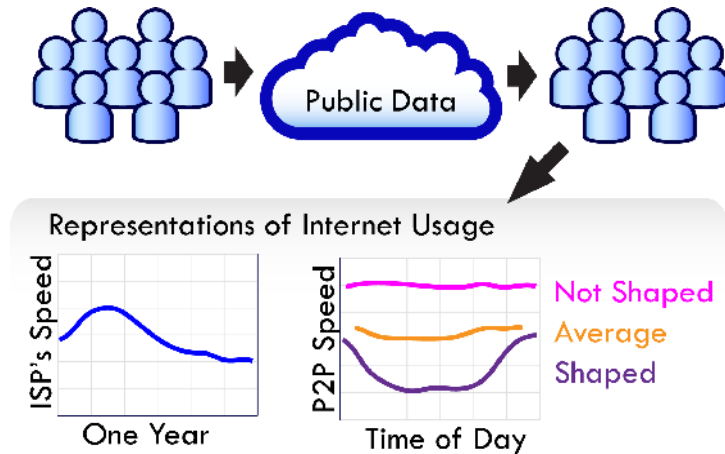
In another example, IXMaps, a project of the New Transparency Project at the Faculty of Information at the University of Toronto, seeks to identify how our information moves across the internet and whether it passes through known points of government surveillance.[18] Concern over internet surveillance arose after a leak revealed the National Security Association and AT&T partnered to install secret rooms in many of the major traffic aggregation hubs on the internet. With the leak came the locations of some of the major surveillance hubs. IXMaps allows users to contribute their traffic routes to reveal whether a users' communication passes these sites or to potentially identify other sites. The research project, in other words, reveals where surveillance might take place on the internet. Crowdsourcing provides a way for the public to participate in this research.

These projects crowdsourced by developing software which allowed users to monitor their traffic, and make conclusions about the nature of their internet connection. In doing so, they mobilized the public as an alternative source of internet usage data. While the term crowdsourcing is new, John Dewey, the American Pragmatist, believed "we lie… in the lap of an immense intelligence".[19] The challenge, as Dewey recognized, was to mobilize this intelligence.

Imagining a Public Research Project

A public internet usage project entails Canadians voluntarily monitoring their usage, pooling this data in a common repository, and sharing it. It might require individuals to install an application on their computer or a website for people to run tests on their internet connection. Importantly, test results would become a public data resource.[20] Some potential uses of this data might include a comparison of ISPs' traffic management practices, accurate bandwidth tests to compare rural and urban internet access, measures of internet speed across the country, the popularity of protocols, and average consumption of bandwidth. Data would not only come from the public, but would be open for public interpretation. Ideally, visualization tools, like Many Eyes[21], would allow the public to explore the data and to better understand their internet usage. Figure 1 illustrates this process of collecting, and then allowing users to create representations of internet usage. It marks a first step toward working through some of the practicalities of such a project.

Figure 1

Development has already begun on software to monitor, collect, and analyze internet usage in the way described above. A consortium of industry and academic researchers[22] created the M-Lab to provide "an open platform for researchers to deploy Internet measurement tools".[23] The website lists a number of different tools to monitor traffic shaping, bandwidth usage, and congestion. The Electronic Freedom Foundation has also begun work on a similar tool, known as the *Switzerland Network Testing Tool*, to test traffic shaping based on its early work with ComCast.[24] Importantly, all these tools exist as open-source projects, so those who are technically adept can scrutinize the code or contribute to its development. Canadian media reform movements should consider partnering with these projects to bring the capabilities of this software into the hands of citizens willing to participate in researching the nature of their internet connection.

The proposed public research project seeks to enhance the democratic policy process. Public data would enable greater participation by both holding Canadian firms accountable and orienting resources and development towards the common good. The hearing on internet traffic management practices showed the need for a public data project as the CRTC places the onus on the complainant to prove a violation of its guidelines. Public data would enable the collection of evidence for these complaints. The ruling on this issue also required ISPs to explain their traffic shaping to the public, but, as Michael Geist, Canada Research Chair of Internet and E-commerce Law at the University of Ottawa, points out, many firms have yet to comply.[25] Finally, many of the technologies driving concerns about network management in Canada, including deep packet inspection (See Parsons in this book), raise important questions about how to manage scarce bandwidth in support of the public good. For example, the First Nation ISP, K-Net, uses traffic shaping to prioritize its community video-conferencing over other traffic.[26] Formulating a similar sense of public good priorities on the wider the internet will prove challenging, but public data would aid this cause.

The project has its risks. The technology to pool information remains in its infancy. The data must always be anonymous and secure. As already noted, this chapter *imagines* a public research project. A project, as Bruno Latour writes, "is a fiction, since at the outset it does not exist".[27] However, this fiction shows clear promise. Public knowledge will enrich how we regulate the internet, how we connect the medium to the public good, and how we hold network owners more accountable.

[1] Shade, L.R., Public Interest Activism in Canadian ICT Policy: Blowin' in the Policy Winds. *Global Media Journal - Canadian Edition* 1 (1), 107-121 (2008).

[2] Wasik, B., *And Then There's This: How Stories Live and Die in Viral Culture*. (Viking, New York, 2009). He writes, "the Internet is revolutionary in how it has democratized not just culture-making, but *culture monitoring*, giving individual creators a profusion of data with which to identify trends surrounding their own work and that of others" (p. 14).

[3] The OECD's webpage for a complete list of their measures, see: http://www.oecd.org/document/54/0,3343,en_2649_34225_38690102_1_1_1_1,00.html.

[4] For the StatsCan News Release summarizes the project, see: http://www.statcan.gc.ca/daily-quotidien/100510/dq100510a-eng.htm.

[5] Entman, R.M., Framing: Toward Clarification of a Fractured Paradigm. Journal of Communication 43 (4), 51-58 (1993).

[6] By way of comparison, framing is also a problem in public opinion polls. As Justin Lewis writes, "opinion polls are often constructed around elite concerns with a narrow framework of response" Lewis, J., *Constructing Public Opinion: How Political Elites Do What They Like and Why We Seem to Go Along with It*. (Columbia University Press, New York, 2001), p. 168.

[7] Canadian Radio-television and Telecommunications Commission Telecom Regulatory Policy CRTC 2009-657 - Review of the Internet Traffic Management Practices of Internet Service Providers (2009). The decision is avaliable online here: http://www.crtc.gc.ca/eng/archive/2009/2009-657.htm.

[8] Bell Canada Comment on Public Notice 2008-19 - Review of the Internet Traffic Management Practices of Internet Service Providers (2009). The archive of comments is available on the CRTC website, here: http://www.crtc.gc.ca/PartVII/eng/2008/8646/c12_200815400.htm.

[9] Canadian Radio-television and Telecommunications Commission Hearings for Internet Traffic Management Practices of Internet Service Providers (2009).

[10] BitTorrent Inc. Comment on Public Notice 2008-19 - Review of the Internet Traffic Management Practices of Internet Service Providers (2009).

[11] For a copy of the report, see: http://www.eff.org/wp/packet-forgery-isps-report-comcast-affair.

[12] See, http://www.wired.com/threatlevel/2008/01/fcc-opens-file/.

[13] For details of the study and the methods, see: http://wiki.vuze.com/w/ISP_Network_Monitor.

[14] For the full report, see: http://wiki.vuze.com/w/Bad_ISPs.

[15] See the coverage from the Canadian Broadcasting Corporation at: http://www.cbc.ca/technology/story/2008/04/22/tech-vuze.html.

[16] Howe, J., Wired 14.06: The Rise of Crowdsourcing Available at http://www.wired.com/wired/archive/14.06/crowds_pr.html, (2006).

[17] For more details, see: http://www.herdict.org/.

[18] For more details about IXMAPS, see its website: http://www.ixmaps.ca/.

[19] Dewey, J., *The Public and its Problems*. (A. Swallow, Denver, 1954 [1927]), p. 219.

[20] Many Internet Speed Tests exists online, see: http://www.dslreports.com/tools. The project differs in that it pools these many testing pooling into a common database, not unlike what Ookla has done with the results of its speed tests to produce their global Net Index, see: http://www.netindex.com/. A public project would differ because from the Ookla Net Index because the data would be open, likely licensed under the GNU Free Documentation License or a Creative Commons license.

[21] Many Eyes is a project hosted by IBM whose "goal is to 'democratize' visualization and to enable a new social kind of data analysis." For more details, see: http://manyeyes.alphaworks.ibm.com/manyeyes/page/About.html.

[22] Measurement Lab founders include New America Foundation's Open Technology Institute, the PlanetLab Consortium, Google Inc. and academic researchers. For a complete list of the project's partners, see http://www.measurementlab.net/who.

[23] Taken from the project description, for more details see the project's website: http://www.measurementlab.net/.

[24] The Switzerland Project contains more details about the project and a download link. See, http://www.eff.org/testyourisp/switzerland.

[25] Geist, M., Geist: ISPs fall short on Net neutrality rules, Available at http://www.thestar.com/news/sciencetech/technology/lawbytes/article/765521--geist-isps-fall-short-on-net-neutrality-rules, (2010).

[26] See: McIver Jr., W., Internet in Media Divides: Communication Rights and the Right to Communicate in Canada, edited by Marc Raboy & Jeremy Shtern (UBC Press, Vancouver, 2010), pp. 145-174 and McKelvey, F. & O'Donnell, S., Out from the Edges: Multi-site Videoconferencing as a Public Sphere in First Nations. Journal of Community Informatics 5 (2) (2009).

[27] Latour, B., *Aramis, or, The Love of Technology*. (Harvard University Press, Cambridge, 1996), p. 23.