2015

Teaching the Network: a Brief Demonstration of the Internet's Structure for Information Literacy Instruction

Robin Camille Davis
CUNY John Jay College

This is an electronic version (postprint) of an article published in *Behavioral and Social Sciences Librarian*. The article is available online at http://www.tandfonline.com/doi/full/10.1080/01639269.2015.1042820 (subscription required).

Full citation:
Davis, Robin. "Teaching the Network: a Brief Demonstration of the Internet's Structure for Information Literacy Instruction." *Behavioral & Social Sciences Librarian* 34.2 (2015): 88-94. Print and web.

Teaching the network: a brief demonstration of the Internet's structure for information literacy instruction

Abstract: A basic understanding of the Internet's physical and operational structure is one element of information literacy. In this article, "traceroute" and "whois" commands are demonstrated as tools that librarians can use to illustrate how the Internet is geographically distributed, how businesses enable and control information sharing, and how to check a source's credibility by determining website ownership. With these tools, students can gain a better understanding of how online information is created, accessed, and affected in ways that may be otherwise invisible.

Global Internet use has existed for as long as today's younger college students have been alive (Shah & Kesan 2007, 100). As they have grown up, so has the infrastructure necessary to support heavy data traffic between millions of machines. The Internet's physical topology has been so skillfully embedded in everyday life that it can seem invisible (Kitchin & Dodge 2011). The ubiquitous technological term "the cloud" belies the masses of cables laid between cities and continents and the many nondescript buildings housing humming servers. When users lack a basic understanding of the physical structure of the Internet and its inherent risks, they rely too much on a "black box," i.e., a system wherein the transformation process from inputs to outputs is not clear. Understanding the Internet's basic physical topology is an essential part of understanding the Internet as a whole and can therefore be considered a part of information literacy.

The Internet's infrastructure is not only physical, but social and economic, too. The cables and routers that power the flow of the Internet are owned by public and private organizations whose interactions and business decisions affect users' experience of the Internet, directly or indirectly. The creation of websites is also bound by social norms that leave traces of metadata, which can prove useful in ascertaining the credibility of a web resource. Librarians invested in providing information literacy instruction should incorporate teaching the physical and operational structure of the Internet into library curricula because its stakeholders, regulation, and limitations have a direct influence on students' online research and everyday information-finding experiences.

This article walks through using *traceroute* and *whois* commands as pedagogical illustrations of the Internet's structure. Note that the information about the workings of the Internet is mainly limited to the United States in this article.

Traceroute

To teach students to begin examining how the network operates, we can turn to an old method: the *traceroute* command. Often used to diagnose network or website issues, traceroute is a simple illustration of how computers connect to websites. This demonstration is also an opportunity to show students that coded technical text that looks obscure can be understandable and refer to real-world things.

To load a webpage, a computer must request its contents, making contact with many routers on its journey. These routers are physical machines that pass the request on to other routers until finally the desired location is reached. Traceroute names every router contacted along the way, ending with the server that hosts the requested website. (In greater detail: traceroute sends out "packets," small chunks of data, with an incrementing TTL, or "time to live" expiration date. These packets successively return to the request's origin with information about the router and how long it took to connect to it [Gibbs 1999].) In a student demonstration, it may be useful to bring back the dated phrase "information superhighway." Just as a car may use a highway that passes through several other cities before arriving at its destination, so a computer's request travels along miles of cables through many heavily-trafficked routers to connect to a website.

On a PC, run a traceroute by opening the Command-Line Prompt (usually under Start » Accessories). Type **`tracert ala.org`** (or any website) and hit enter. On a Mac or in Linux, open Terminal (usually under Applications » Utilities on a Mac, or the desktop application menu in Linux). Type **`traceroute ala.org`** (or any website) and hit enter. The traceroute command will produce a sequential list of the routers' IP addresses as the request travels to ala.org, with timestamps indicating how long it took to connect to each. First-time users of the command line may find the visual display dauntingly sparse, but with close inspection, the information displayed is a textual map.

```
Home-computer:~ home$ traceroute ala.org
traceroute to ala.org (216.80.72.244), 64 hops max, 52 byte packets
1 10.0.1.1 (10.0.1.1) 4.955 ms 4.613 ms 3.022 ms
2 * 10.240.161.17 (10.240.161.17) 250.738 ms 10.490 ms
3 67.59.228.205 (67.59.228.205) 9.853 ms 11.268 ms 9.368 ms
4 rtr3-ge1-1.mhe.hcvlny.cv.net (67.83.222.1) 12.478 ms
5 rtr101.wan.hcvlny.cv.net (65.19.99.125) 12.359 ms
6 64.15.3.234 (64.15.3.234) 13.950 ms 11.659 ms 13.933 ms
7 paix.nyw.ny.rcn.net (198.32.118.23) 12.947 ms 14.693 ms 13.711 ms
8 bdle4.core1.nyw.ny.rcn.net (207.172.15.84) 13.481 ms
9 tge0-1-0-6.core2.sbo.ma.rcn.net (207.172.19.12) 41.683 ms
10 tge0-0-0-1.core2.chgo.il.rcn.net (207.172.19.233) 44.253 ms 47.482
   ms
11 bdle2.aggr2.chgo.il.rcn.net (207.172.15.217) 39.979 ms
12 www.ala.org (216.80.72.244) 39.214 ms 38.457 ms 42.890 ms
```

Figure 1. Traceroute result for ala.org from a computer in New York City

The first item listed in each line is the host name, the semi-human-readable form of the IP address, which follows in parentheses. The time it took to connect is listed in milliseconds. In the example given in Fig. 1, subtle geographic information is provided in the traceroute result. From a computer in New York City, the request to load ala.org travels first through the local (in-building) network, indicated by IP addresses beginning with 10. Then, starting in line 4, the request travels to Hicksville, NY (*hcvlny*). This may not be apparent at first glance, but Googling all or part of a host name can bring up information about it. In line 7, the request travels to a router that is probably in western New York State (*nyw.ny*). Guessing at a similar naming convention used in line 9, the request now travels to Massachusetts, possibly South Boston (*sbo.ma*) and on to Chicago in line 10 (*chgo.il*), where the American Library Association and its servers are based.

Not all traceroute commands will produce a transparent or guessable geographic path. Before demonstrating, it is a good idea to have already performed a traceroute with a salient example, building in time to run traceroutes on websites that students suggest, too. Running traceroutes on known foreign websites demonstrate how global connections are made, and particularly how they take more time to connect, since they are physically farther away. The "map" that traceroute produces illustrates the very physical character of the Internet, demonstrating how online information has a physical "home," and that information sources tend to cluster in large, well-connected cities in well-connected countries.

Traceroute also provides a glimpse at the operational structure of the Internet. Most host names indicate ownership. In Fig. 1, lines 4–6 indicate that the request made contact with routers with a hostname ending in *cv.net*, indicating that they are owned by Cablevision Systems Corporation, a publicly traded company. Starting in line 7, the routers' hostnames end in *rcn.net*, indicating RCN Telecom Services, a privately held company. Other New York City-based computers' requests to load other Chicago-based websites may run through these same companies' routers, too, as they form part of the "Internet backbone." The backbone is composed of many telecommunication companies' networks, which connect to each other through "peering" and to smaller local networks—in brief, allowing computers to connect to websites (Yurcik 2002). Because the network of networks is so heavily dominated by large businesses, the Internet experience is prone to changes due to business deals and regulatory measures. For example, in February 2015, Federal Communications Commission (FCC) voted to uphold net neutrality, protecting heavy Internet users from punishingly slow speeds at the hands of their Internet service providers (ISPs). One contentious part of the FCC's debate was uncovering a paid peering deal between Netflix and ISPs (Brodkin 2014). The FCC's opinion shifted throughout the debate, largely due to public outcry as general Internet users learned how they would be affected by a policy change. As consumers of the Internet, students should have some understanding of how their experience of the Internet is controlled by many different companies whose business decisions and government regulations affect them. Digitally literate students are knowledgeable consumers and constituents who take part in conversations about how information is shared and controlled.

Whois

Every website must be registered with the Internet Corporation for Assigned Names and Numbers (ICANN). When registering a domain name, like ala.org, the individual or

organization must provide information including the registrant's name, address, phone number, email address, administrative contact, and technical contact (Internet Corporation for Assigned Names and Numbers n.d.). This information is publicly viewable with the *whois* command.

On a PC, open the Command-Line Prompt. On a Mac or Linux, open Terminal. Type **whois ala.org** (or any website) and hit enter. The resulting display will be similar to Fig. 2.

```
Home-computer:~ home$ whois ala.org
Domain Name:ALA.ORG
Domain ID: D931483-LROR
Creation Date: 1994-01-18T05:00:00Z
Updated Date: 2014-09-18T22:25:52Z
Registry Expiry Date: 2017-01-19T05:00:00Z
Sponsoring Registrar:Network Solutions, LLC (R63-LROR)
Sponsoring Registrar IANA ID: 2
WHOIS Server:
Referral URL:
Domain Status: clientTransferProhibited --
http://www.icann.org/epp#clientTransferProhibited
Registrant ID:23304167-NSI
Registrant Name:American Library Association
Registrant Organization:American Library Association
Registrant Street: 50 East Huron Street
Registrant City:Chicago
Registrant State/Province:IL
Registrant Postal Code:60611
Registrant Country:US
Registrant Phone:+1.9999999999
Registrant Phone Ext:
Registrant Fax: +1.9999999999
Registrant Fax Ext:
Registrant Email:svanyek@ALA.ORG
. . .
```

Figure 2. Beginning of the whois record for ala.org.

It is important to note that this command is not foolproof: contact information is not verified and could be false; some web hosting services offer to mask site owners' identities by providing non-names like "A Happy Customer" and proxy addresses (Dreamhost 2011); and though ICANN asks site owners to update their information annually, the whois record may not be up to date.

The whois command is pedagogically useful in an information literacy context. The ACRL "Framework for Information Literacy for Higher Education" identifies "Authority Is Constructed and Contextual" as a key concept (Association of College & Research Libraries 2015). Under this concept, one knowledge practice is to "[u]se research tools and indicators of authority to determine the credibility of sources, understanding the elements that might temper this credibility." Though whois does not offer a full description of site ownership or authorship, it can confirm hunches when a website or its URL does not seem to match what it purports or appears to be. For students doing online research, it can be advantageous to check the whois record of a website that seems to present dubious information or that does not otherwise provide ownership or authorship attribution. In

everyday life, checking URLs linked in suspicious emails with a whois query can save a user from a phishing scam. Less nefarious but still deceptive are websites that look exactly like brand-name retailers and purport to be sale sites, but which are unauthorized sellers that may be selling fakes. As students grow more information-literate, they hone gut feelings about fishy sources. A whois query can be one tool to confirm or allay suspicions.

Demonstrating whois queries is also a way to incorporate the "Information Creation as a Process" concept from the Framework into information literacy instruction. Students who have never registered their own domain name may be surprised to learn about the process. The whois record contains an expiration date, set when the domain name is purchased. This item of information touches on two points: the inherent cost of registering a website (not to mention maintaining it) and the changing nature of the Internet. If the domain name registration is not renewed by its expiration date, the website disappears. Sometimes websites fall away organically, when a website's owner lets it die. Other times, the failure to renew may be a high-impact error, as when the entire *New York Magazine* website was rendered inaccessible when a staff member forgot to renew (Sebastian 2014). That websites must actively maintain their online status is one distinct characteristic of online information. Students should be aware that online information is subject to unreliable access due to a variety of factors, including domain name registration changes among many others.

Conclusion

Traceroute and whois are two basic command-line tools that can demonstrate to students the physical and operational structure of the Internet. Traceroute is one textual representation of the Internet's physical topology, tracing the route of a computer's request to load a given webpage. Traceroute also reveals some information about the businesses that enable and control Internet connections. Whois displays registration information for a given website, revealing some insight into its credibility and into the changing nature of online information. With these tools, students can gain a better understanding of how online information is created, accessed, and affected in ways that may be otherwise invisible.

Bibliography

Association of College & Research Libraries. “Framework for Information Literacy for Higher Education,” 2015. http://www.ala.org/acrl/standards/ilframework.

Brodkin, J. “FCC Gets Comcast, Verizon to Reveal Netflix’s Paid Peering Deals.” *Ars Technica*, June 13, 2014. http://arstechnica.com/tech-policy/2014/06/fcc-gets-comcast-verizon-to-reveal-netflixs-paid-peering-deals/.

Dreamhost. “Domain WHOIS Privacy Service,” 2011. http://wiki.dreamhost.com/Domain_WHOIS_Privacy_Service.

Gibbs, M. “The Inner Workings of Traceroute.” *Network World*, July 12, 1999, 42.

Internet Corporation for Assigned Names and Numbers. “About WHOIS,” n.d. https://whois.icann.org/en/about-whois.

Kitchin, R., and M. Dodge. *Code/Space*. Cambridge, MA: MIT Press, 2011.

Sebastian, M. “Glitch Crashes New York Mag Website, Gives Twitter Users Something to Joke About.” *Ad Age*, April 23, 2014. Accessed February 20, 2015. http://adage.com/article/media/york-magazine-s-website-felled-technical-glitch/292806/.

Shah, R.C., and J.P. Kesan. “The Privatization of the Internet’s Backbone Network.” *Journal of Broadcasting & Electronic Media* 51, no. 1 (March 2007): 93–109. doi:10.1080/08838150701308077.

Yurcik, W. “Internet: Backbone.” In *Computer Sciences*, edited by Roger R. Flynn, 4:142–46. Electronic Universe. New York: Macmillan Reference USA, 2002.