KEEPING THE INTERNET
OPEN • INNOVATIVE • FREE

www.cdt.org

CENTER FOR DEMOCRACY
& TECHNOLOGY

1634 I Street, NW
Suite 1100
Washington, DC 20006

P +1-202-637-9800
F +1-202-637-0968
E info@cdt.org

Comments of the Center for Democracy & Technology on The Federal Trade Commission's "The Big Picture: Comprehensive Online Data Collection Workshop"

March 8, 2013

The Center for Democracy & Technology (CDT) welcomes the opportunity to provide comments on comprehensive data collection and commends the FTC for holding a workshop on this pressing topic. CDT has been concerned for years about the implications of comprehensive data collection, having testified against the use of deep packet inspection for behavioral advertising without clear, affirmative consent,[1] and supported Representative Rush's BEST PRACTICES Act legislation which included heightened obligations for the collection of "all of substantially all" of a user's online activity.[2] We were one of the first organizations to object to NebuAd's ISP-level monitoring program when it was unveiled in 2008,[3] and we were heartened when ISPs eventually rejected using NebuAd's technology to monitor all their customers' communications.[4] However, in recent years, we have seen companies start to engage in behaviors very similar to NebuAd's. We urge the FTC to recognize the special privacy threats associated with comprehensive collection and to call for stronger privacy protections for the creation of comprehensive databases.

Why Comprehensive Matters

Comprehensive (or sometimes, near comprehensive) data collection is special for two reasons. First, and most obviously, is the expanded scope of information that a monitoring party has about the individual. Users have an inherent privacy

---

[1] Center for Democracy & Technology, Statement of Alissa Cooper before the Committee on Energy and Commerce, "Comments of the Center for Democracy and Technology, Consumers Action, and Privacy Activism: In regards to the FTC Staff Statement, 'Online Behavioral Advertising: Moving the Discussion Forward to Possible Self-Regulatory Principles," April 11, 2008, https://www.cdt.org/privacy/20080411bt_comments.pdf.

[2] Center for Democracy & Technology, Statement of Leslie Harris before the Committee on Energy and Commerce, "The Best Practices Act of 2010 and Other Federal Privacy Legislation," July 22, 2010 < https://www.cdt.org/files/pdfs/CDT_privacy_bill_testimony.pdf.

[3] Alissa Cooper, An Overview of the Federal Wiretap Act, Electronic Communications Privacy Act, and State Two-Party Consent Laws of Relevance to the NebuAd System and Other Uses or Internet Traffic Content from ISPs for Behavioral Advertising, July 8, 2008, https://www.cdt.org/privacy/20080708ISPtraffic.pdf; .Ryan Paul, *NebuAd's "breakthrough opt-out" approach: legal or no?*, ARSTECHNICA, July 9, 2008, http://arstechnica.com/uncategorized/2008/07/nebuads-breakthrough-opt-opt-approach-legal-or-no/.

[4] Alissa Cooper, Backing Down on Behavioral Advertising, October 13, 2008, Center for Democracy & Technology Blog, https://www.cdt.org/blogs/alissa-cooper/backing-down-behavioral-advertising.

interest (if not necessarily a legal *right*) in the information that is gathered about them, and a user is going to care considerably more if someone knows 10,000 facts about her instead of just one. Recent arguments to require an articulation of privacy harm, or to consider only use-based privacy rules, ignore the interest that a person has in not being persistently monitored. Even when data is collected merely for limited purposes, consumers could reasonably worry that their data could later be used for new, unexpected and unwanted purposes, accessed and misused by a rogue employee, breached by hackers, unwittingly exposed, or accessed by the government without robust legal process.[5]

These concerns are multiplied in the context of comprehensive collection, as these data sets pose a considerably more alluring target to those who would access it illegitimately, with far greater exposure of personal information. The knowledge that consumer behavior is being monitored and retained (and potentially shared, accessed, or lost) can have a chilling effect on free expression, as well as the adoption of new technologies and services.[6] Inability to control the collection of information represents an intrinsic limitation to user autonomy: That is, in order to have a healthy democratic society, people need their own "safe spaces" in which to make mistakes, test out theories and do other activities that they may not engage in if being surveilled.[7]

The other problem with comprehensive data collection is that it by definition must be collected by a platform or intermediary — since no one site or content provider can get a comprehensive view into all of a user's online activity — and thus is likely to be unexpected and out of context. Traditional first party data collection and tracking is relatively intuitive and understandable: When a user goes to a website like NYTimes.com or Amazon.com, it is not altogether surprising that those sites are able to keep state on a user over time to count read articles or suggest new products. The user understands that she was communicating with that entity.

However, third party collection is more unexpected, and potentially less desirable. We do not typically think of or want our communications with others being monitored, even by the intermediaries we use to communicate; for this reason, Congress passed the Wiretap Act to place strong limitations on the ability of parties to intercept and monitor personal communications.[8] Courts have found that online behavioral advertising is permissible under the Wiretap Act because the sites we visit consent to the collection.[9] However, those companies still recognize the value in messaging to users that cross-site collection is happening, but their substantial efforts have met with mixed results. Even after attaching the DAA AdChoices to online ads trillions of times,[10] most users do not

---

[5] The Center for Democracy & Technology, Statement of Justin Brookman before the Committee on Energy and Commerce, "Hearing on 'Balancing Privacy and Innovation: Does the President's Proposal Tip the Scales?'", March 29, 2012, https://www.cdt.org/files/pdfs/Justin-Brookman-privacy-testimony.pdf.

[6] *Id.*

[7] Julie E. Cohen, *What is Privacy For?*, 126 Harv. L. Rev. __ (forthcoming 2013).

[8] 18 U.S.C. §§ 2510-2522.

[9] *In re DoubleClick Inc. Privacy Litigation*, 154 F. Supp. 2d 497 (S.D.N.Y. 2001).

[10] Press Release, *DAA Announces New Managing Director*, Digital Advertising Alliance, August 21, 2012, http://www.marketwatch.com/story/daa-announces-new-managing-director-2012-08-21.

understand how behavioral advertising works or how to control it.[11] Third-party online collection is hard to message and explain.

Given the relative privacy concerns, it is somewhat surprising that online behavioral advertising and "Do Not Track" have gotten the lion's share of popular attention and regulatory scrutiny in recent years instead of comprehensive collection (at least post-NebuAd). From a privacy perspective, comprehensive collection poses a considerably greater threat, given the scope of the data in question as well as the fact that the data is more likely to be collected on a real name basis. While users are more likely to have a direct relationship with the monitoring party in comprehensive collection scenarios, as noted above, platform-level collection and monitoring is often not contextually evident, and users would not expect or want it anymore than they would expect the postal service to examine the contents of written communications. Perhaps behavioral advertising has received such outsized attention because understanding of those information collection practices has matured over the course of several years, whereas comprehensive tracking is relatively recent. For this reason, we believe the FTC should issue clear guidance now, rather than wait for the inevitable recognition and popular outcry, and try to retrofit existing business models with after-the-fact privacy protection and control (as we have seen happen in the behavioral advertising space).

Comprehensive Data Collection Merits More Stringent Privacy Protections

For years, CDT has argued that comprehensive data collection should by and large only be done on an affirmative, opt-in basis. And because such collection is typically done out of context, messaging and obtaining informed consent is often going to be very challenging. The FTC has previously taken action against companies that engage in comprehensive collection without a clear, disclosed need or user consent,[12] and we hope the FTC will continue to be aggressive to curtail excessive comprehensive collection practices.

We continue to believe that comprehensive data collection should only be done an affirmative opt-in basis, unless the platform can demonstrate a compelling need for the information. Even then, the data collectors should provide prominent and clear transparency about the collection and adhere to narrow retention periods, keeping the data only for as long as necessary to achieve the compelling need. For some of the bourgeoning comprehensive data collection programs that are appearing today, it is not clear this standard is being met.

Opt-in consent to value-add features

The most obvious and legitimate way to obtain permission for comprehensive collection would be to make a value proposition to consumers and persuade them of the benefits

[11] Pedro Leon et al., *What Do Online Behavioral Advertising Disclosures Communicate to Users?*, Carnegie Mellon Cylab, April 13, 2012, <http://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab12008.pdf>; Wendy Davis, *Study: Web Users Don't See AdChoices Icon*, MEDIAPOST, November 13, 2012, http://www.mediapost.com/publications/article/187164/study-web-users-dont-see-adchoices-icon.html#axzz2N5ZjP53i.

[12] Federal Trade Commission, Sears Settles FTC Charges Regarding Tracking Software, June 4, 2009 http://ftc.gov/opa/2009/06/sears.shtm.

of comprehensive collection. For example, Google's Chrome browser typically does not transmit comprehensive information about a user's browser communications to Google (though comprehensive information is stored client-side in the user's browser history file and cache). However, users have the ability to affirmatively "Sign in to Chrome" if they want their browser history, applications, and bookmarks stored in the cloud and synced across various devices.[13] This type of comprehensive collection seems fairly understandable and straightforward, as the user makes a clear choice to turn on this collection feature. Google also offers products like Screenwise Select, where a user receives cash payments in exchange for agreeing to install a Google modem at home to let the company monitor all the user's web traffic for research purposes.[14] Again, this collection is conducted pursuant to an intuitive, above-board transaction, and we don't need to second-guess the individual's freely given choice.

Comprehensive collective necessary for functionality

Other products may have comprehensive data collection built into their functionality as a matter of necessity. Proxy browsers — web browsers that collect URI requests and render web sites on company servers before displaying them on the user's client — allow less powerful devices to take advantage of the full functionality of the web. They often also decrease customer's data usage, an increasingly important issue to consumers as ISPs have pulled back on unlimited data plans. In many cases, a company may not necessarily have to use a proxy service to generate web content, though proxy service does allow companies to sell devices to users at a lower price point, as sophisticated client-side computing is less necessary.

As one example, in September 2011, Amazon released the Silk browser, which by default passes web requests to Amazon's cloud service, where Amazon renders the page more efficiently than it could on the device.[15] The Silk browser temporarily logs URLs for the pages it serves and originating IP addresses, and keeps this information for up to 30 days. Users have to ability to opt out of connecting the browser to Amazon's servers, and by default, encrypted communications are not proxyed, but are instead generated by the Silk browser on the client.[16]

We have concerns that proxy browsers violate the fundamental end-to-end principle of the web — with concomitant loss of autonomy by the user and requisite monitoring of communications — but on the other hand, we recognize the value they provide for users. Because comprehensive data collection by the manufacturer of a device is not intuitive, companies that deploy proxy browsers have a special responsibility to communicate to users that the data collection is occurring. We appreciate that Amazon has set a

---

[13] Google, Sign in to Chrome, https://www.google.com/intl/en/chrome/browser/signin.html.

[14] Casey Johnston, "Google paying users to track 100% of their Web usage via little black box," ARSTECHNICA, February 8, 2012, http://arstechnica.com/gadgets/2012/02/google-paying-users-to-track-100-of-their-web-usage-via-little-black-box/.

[15] Aaron Brauer-Rieke, "Amazon's Silk Browser Awaits Privacy Assurances," Center for Democracy & Technology blog, September 29, 2011 https://www.cdt.org/blogs/aaron-brauer-rieke/1910amazon%E2%80%99s-silk-browser-awaits-privacy-assurances.

[16] Amazon, *Amazon Silk Terms and Conditions*, http://www.amazon.com/gp/help/customer/display.html/?nodeId=200775270.

relatively short data retention period before anonymization, and makes limited secondary usage of user communications, but it is not entirely clear that logging is necessary or that 30 days' retention is intuitive to users. However, this affirmative limitation is considerably better than what is provided by other proxy browsers (such as Opera and Blackberry) who so far as we can tell make no affirmative limitation on retention for data received through its proxy services.

Social widgets

Cross-site data collection by social widgets is another example of not-quite-comprehensive data collection by a platform that merits special concern. Social networks such as Facebook, Twitter, and Google Plus all allow websites to embed code that allows logged-in users to share directly content from around the web to their social networking circles. Because the social networks generate the sharing buttons themselves (the other websites embed code that calls to social networks' servers to display the buttons), they can tell that their logged-in user has visited a particular page even before the user clicks on the button. Due to the wide range of deployment of these sorts of sharing buttons, the scope of web usage obtained through these widgets is extensive if not comprehensive.

The Facebook "Like" button received considerable attention[17] when it was originally deployed across the web, as advocates realized that Facebook now had a fairly pervasive view of users' off-Facebook surfing — in addition to the detailed information that Facebook already had about its users. In response to this concern, many of these companies have made affirmative representations to delete or anonymize the data with a relatively narrow period of time (90 days for Facebook,[18] "usually" 2 weeks for Google,[19] and 17 days for Twitter[20]). Moreover, each has promised to use the data for only limited purposes, including in the case of Facebook and Google to forego personalization based on this data. Again, it is questionable whether logging this data is necessary or understood by users, and at the very least there should be easy tools for users to turn off the data collection if they decide that personalized content is not worth the privacy invasion (Twitter does halt collection of the data in response to a Do Not Track signal[21]).

---

[17] Declan McCullagh, *Facebook 'Like' button draws privacy scrutiny*, CNET, June 2, 2010, http://news.cnet.com/8301-13578_3-20006532-38.html; Riva Richmond, *As 'Like' Button Spread, So Do Facebook's Tentacles*, NEW YORK TIMES, September 27, 2011, http://bits.blogs.nytimes.com/2011/09/27/as-like-buttons-spread-so-do-facebooks-tentacles/.

[18] Facebook, "What information does Facebook get when I visit a site with the Like button or another social plugin?," http://www.facebook.com/help/186325668085084/?q=pluginds&sid=0CnsdsFl6S0w9XwnZ.

[19] Google, "How the +1 button respects your privacy" http://support.google.com/plus/answer/1319578?hl=en.

[20] Twitter, "Twitter Privacy Policy," https://twitter.com/privacy.

[21] Twiter, "Twitter Supports 'Do Not Track,'" https://support.twitter.com/articles/20169453-twitter-supports-do-not-track.

ISP-level comprehensive collection

Finally, some ISPs such as Verizon Wireless and Sprint have begun monitoring network communications for a variety of reasons, including market research[22] and behavioral advertising.[23] To date, both are only engaging in behavioral advertising on an opt-in basis. Both also offer an opt out for the use of customer data in market research reports, though it is not clear that this opt out extends to the collection of personal information as well. Neither is clear about the scope of data retention for these (or other purposes), though Verizon does say that for users who opt in to the behavioral marketing program, data is retained for up to three years.[24]

Absent a clearly disclosed compelling need, an ISP should not monitor its paying customers' network communications without affirmative opt-in consent. It is conceivable that these (and other) ISPs have a security need to inspect their customers' traffic; however, that case has not been publicly made, and it is not clear that ordinary users understand that their communications are being monitored and stored. Moreover, even if an ISP collects customer data for one legitimate reason, that does not necessarily justify all other uses — including relatively benign usage such as market research that will not change the experience of the customer. Allowance of secondary usage can have the perverse incentive of causing a company to exaggerate the retention period that is necessary for a purpose such as security. Such a compromise — allowing market research on data when collected for another legitimate purpose — was recently rejected in the W3C negotiations on Do Not Track for this very reason.

CDT strongly believes that ISP-level monitoring and retention — and other forms of comprehensive data collection — merit the FTC's close attention in the months and years ahead. We urge the FTC to find that comprehensive data collection should only be done an affirmative opt-in basis, absent a prominent disclosed compelling need for the information. Even then, the data collectors should provide prominent and clear transparency about the collection and adhere to narrow retention periods, keeping the data only for as long as necessary to achieve the compelling need.

For more information, contact Meredith Whipple, mwhipple@cdt.org, 202.637.9800.

---

[22] Verizon, "Precision Market Insights," http://business.verizonwireless.com/content/b2b/en/precision/precision-market-insights.html; Sprint, "Sprint Mobile Advertising and Reporting & Analytics Programs," http://newsroom.sprint.com/article_display.cfm?article_id=1623.

[23] Verizon, "Verizon Selects FAQs," http://support.verizonwireless.com/faqs/Account%20Management/verizon-selects.html; Sprint, "Sprint Mobile Advertising and Reporting & Analytics Programs," http://newsroom.sprint.com/article_display.cfm?article_id=1623.

[24] Verizon, "Verizon Selects FAQ," http://support.verizonwireless.com/faqs/Account%20Management/verizon-selects.html.