Casting light on the Internet's shadows (and shadowing)

November 6 2014, by Catherine Shen

The specter of a faceless system collecting data from Web users and compiling personal profiles has raised alarms among privacy advocates worldwide. Arvind Narayanan, an assistant professor of computer science at Princeton University, founded the Web Transparency and Accountability Project (WebTAP) at the Center for Information Technology Policy (CITP) to address difficult questions related to Internet privacy. How can regular Web users protect themselves from third-party trackers? What can policymakers can do? Could greater transparency and awareness benefit both businesses and everyday Web users?

Narayanan joined Solon Barocas, a postdoctoral research associate at CITP, to provide perspectives on the ethical and social implications of online tracking.

Question: How did WebTAP come to fruition?

Barocas: Two years ago I mentioned to Arvind my idea of capturing and analyzing individually targeted political advertisements online. Arvind thought a more general and ambitious system could be built, one that could reverse engineer all sorts of targeting and personalization schemes on the Web. Arvind and his graduate students began to build out the necessary infrastructure, and WebTAP took concrete form. Meanwhile, I finished my dissertation and joined CITP, where I've been able to bring to bear my expertise on the ethics of data mining on the Web.

Q: Differential treatment of users stemming from algorithms and data use is emerging as a significant concern. How can that harm a regular Web user?

Barocas: A good example comes from an important paper by Latanya Sweeney at Harvard, where she found Google queries for black-sounding names were more likely to return contextual advertisements for arrest records than those for white-sounding names. Sweeney confirmed the companies paying for these ads had not set out to focus on black-sounding names; rather, the fact that black-sounding names were more likely to trigger such advertisements seemed to be an artifact of the algorithm that Google employs to determine which advertisements to deliver alongside the results for certain queries. This aspect of the process could result in the differential delivery of advertisements that reflect the kinds of prejudice held by those exposed to the advertisements.

The shorter answer regarding harm is that any advertisement that suggests an arrest record is likely to taint the impression decision-makers have about the person whose name they've Googled. Think, for example, of an employer who Googles a job applicant.

Narayanan: In addition to ads, search results, news feeds, product recommendations, and, in some cases, prices are all known to be tailored to the user. Since these are things the user is actively looking for, our online experience is shaped by the effect of our past actions, as judged by a data-mining system.

Q: With social media platforms like Facebook having such prevalence on the Web, what consequences did you find of browsing a page with a Facebook "like"

button? What consequences are there when a user clicks or doesn't click the button?

Barocas: There are two different issues to note here. First, each website that includes a Facebook like button on its pages lets Facebook know that you've visited that website, whether or not you've clicked on the button. The information that it gleans from these kinds of passive observations can affect how it personalizes your experience on its website, but also the kinds of advertisements you'll see. Second, Facebook can also learn a lot from those occasions where you affirmatively click the like button, and not just what you might imagine. Researchers have shown that all sorts of inferences can be drawn about you simply based on what you like. The implications of liking something for your experience on Facebook and the rest of the Web are entirely unclear. Researchers are working to measure these effects on the newsfeed, and many other platforms could be brought to bear on these questions, too.

Q: Should we treat those who blatantly prey on Web users differently from those who are ignorant of, for example, third-party trackers on their sites?

Narayanan: WebTAP's thesis is that more transparency is always better in the long run, regardless of a company's motives. For us, transparency isn't just about coming clean to consumers, although that's important, of course. Publishers, the press, and regulators all benefit from having access to information about tracking and personalization. We've found that even if only a fraction of consumers makes choices based on privacy, it can exert a significant pressure on companies to change their practices.

Q: What are some surprising discoveries you've had through the project?

Narayanan: We did a study of canvas fingerprinting, a sneaky type of online tracking that has seen remarkably rapid uptake. The technology went from being introduced in a research paper to being the subject of an open-source project to being adopted by a few small players to being deployed by AddThis, a widget that gets over a hundred billion monthly page views according to Comscore.

In the same study, we also found that cookie syncing is rampant—a technique that allows different tracking companies to match their pseudonymous IDs of you with each other. Once two companies sync their cookies of you, they're in a position to exchange data about you behind the scenes, out of reach of our transparency tools.

In another surprise, we found that an eavesdropper on the network who uses advertising cookies to track individuals—as the National Security Agency has been revealed to do—can reconstruct essentially the entirety of their online browsing activity, and link it to their real-world identity.

Q: Despite all the time and money spent on developing online ads, many Web users ignore them. Why is that, and how do the continued efforts to improve advertising affect the consumer?

Barocas: Online advertisers are in the business of making marginal improvements in the so-called click-through rate, the percent of people exposed to an ad that click on it. The vast majority of online advertisements never get any attention; commonly, click-through rates are less than 1 percent. The only reason that most advertisers can survive on such low click-through rates is that they're often serving many millions—if not billions—of ads every day.

A 1 percent click-through rate for this enormous number can still generate a significant sum of money.

Presumably, the current click-through rates reflect some kind of informed trade-off between the cost of being able to target more effectively and the expected benefit, but individual players within the industry will continue to compete with one another to find new ways to generate greater return on investment. The users are likely to suffer in this endless arms race because they'll be subject to ever-more elaborate tracking and sophisticated profiling as a result.

Q: What is the next step for you in terms of research and the future of policymaking?

Narayanan: Research on Web transparency is at a critical juncture. One of the most important trends in the industry has been the merging of online and offline tracking. Companies can now use their customer databases from physical stores to target those same customers online, or personalize in-store deals and coupons based on a variety of data including consumers' online activities. This is a worrisome development, because it opens further avenues of profiling and manipulation. We need more transparency about the types of targeting that are happening. Can researchers keep up? It's easy to create a bot that will go look at prices online; we can't create one that will go shopping in stores, so we'll need to get creative.

Meanwhile, for the existing body of research to achieve its full impact, we'll need stronger ties with policymakers, regulators and self-regulators, and enforcement agencies. Our recent "Web Privacy and Transparency" conference was a first step, but we're always exploring avenues for closer collaboration.

Provided by Princeton University

Citation: Casting light on the Internet's shadows (and shadowing) (2014, November 6) retrieved 20 March 2024 from https://phys.org/news/2014-11-internet-shadows-shadowing.html