

Unearthing trackers of the past: Computer scientists reveal history of third-party web tracking

15 August 2016, by James Urton



Credit: Flickr/Creative Commons

For over two decades, consumers have used the internet to research, shop, make friends, find dates and learn about the world with the click of a mouse or a few keystrokes. But as we've surfed and tweeted, third-party watchers have also been watching—and learning—about us.

When you open a website, your browser doesn't just talk to the site you've intended to visit. The site may contain "third parties"—other embedded websites that your browser also talks to such as advertisers, website analytics engines or [social media](#) widgets—which can observe your browsing behavior. Often these companies use this information for innocent—albeit sometimes intrusive—applications like targeted advertisements or personalized content. But third party web trackers raise questions about user privacy, as they can identify users as they visit multiple sites, pick up a person's trail and potentially construct a comprehensive profile based on web behavior.

At the [USENIX Security Conference](#) in Austin, Texas, a team of University of Washington researchers on Aug. 12 presented the [first-ever comprehensive analysis of third-party web tracking](#) across three decades and a new tool, TrackingExcavator, which they developed to extract and analyze [tracking](#) behaviors on a given web page. They saw a four-fold increase in third-party tracking on top sites from 1996 to 2016, and mapped the growing complexity of trackers stretching back decades.

"Third-party tracking started quite early in the history of the web," said Adam Lerner, a graduate student in the UW Department of Computer Science & Engineering who presented the team's findings at the conference. "People are becoming more concerned about the potential impact of third-party web tracking, but we lacked a comprehensive history of how trackers—and the types of information they collect—have evolved over time."

Lerner and fellow doctoral student Anna Kornfeld Simpson set out to fill the gaps in our understanding of tracking, working with computer science and engineering assistant professor Franziska Roesner and associate professor Tadayoshi Kohno of the UW Security and Privacy Laboratory.

Roesner and Kohno [previously studied third-party web tracking techniques](#), including developing an early taxonomy of the basic approaches that many cookie-based trackers employ.

"Tracking behavior ranges from something 'forced,' like a pop-up window, to something more 'vanilla' like a third-party cookie that tracks the user," said Kohno. "Until now, we didn't have the tools to understand how these approaches have changed since the earliest days of the web. Now we can see

how the quantity and variety of trackers has grown, and how some approaches have fallen out of favor while others are on the rise."

The project was no small feat, since no one has been systematically collecting information about tracking over time. To overcome this limitation, TrackingExcavator gathers data from an extensive, open-access archive of websites known as the [Wayback Machine](#), which preserves website content as far back as 1996.

"Reconstructing tracking behavior from the Wayback Machine is difficult because it was designed to archive web content, not tracking techniques," said Kornfeld Simpson. "We had to develop techniques to extract tracking information from the archive. For example, we collected tracking cookies from archived HTTP headers and Javascript and then simulated the browser's cookie storage behaviors to detect tracking behavior."

This complex reconstruction occupied much of the team's time over the past year, but the end result is a historical overview of third-party tracking trends for top internet sites from 1996 to 2016. They quantified the increase of third-party web tracking and illustrated the emergence of different tracking techniques over time.

In 1996, the average number of third-party requests on top websites was less than one. Ten years later, that number rose to about 1.5. Today, the average top website has an average of at least four third-party trackers looking at user activity. The team stresses that these numbers are likely underestimates, since not all websites are fully archived.

They also found that today individual trackers cover a much larger fraction of the web. Before 2003, no single tracker could observe browsing behavior on more than about 5 percent of the most popular sites. That number increased to 10 percent by 2007. Today, many popular trackers have expanded their coverage to at least 20 percent of sites, while one third party—Google Analytics—is on over a third of the most popular sites. These findings are important to understanding the effects of tracking on privacy, since tracking users on more

sites allows trackers to develop a more detailed and intimate picture of their behavior.

This 20-year historical perspective paints a clear picture of how third-party tracking has evolved with the rise and fall of different techniques, advances in technology, and our increasing reliance on the web in our lives. In general, third parties are watching and collecting information. How we may feel about that remains to be seen.

"Without contextualizing today's tracking behaviors in the history of the web, we don't know whether users should have growing concerns about their privacy or whether privacy advocates are crying wolf. Moreover, we can't assess whether media outcries, policy discussions or changing browser defaults are having an effect," said Roesner. "Our work gives us the tools to answer these questions. And our findings suggest that web tracking should remain an area of concern for privacy advocates."

Provided by University of Washington

APA citation: Unearthing trackers of the past: Computer scientists reveal history of third-party web tracking (2016, August 15) retrieved 4 December 2020 from <https://techxplore.com/news/2016-08-unearthing-trackers-scientists-reveal-history.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.