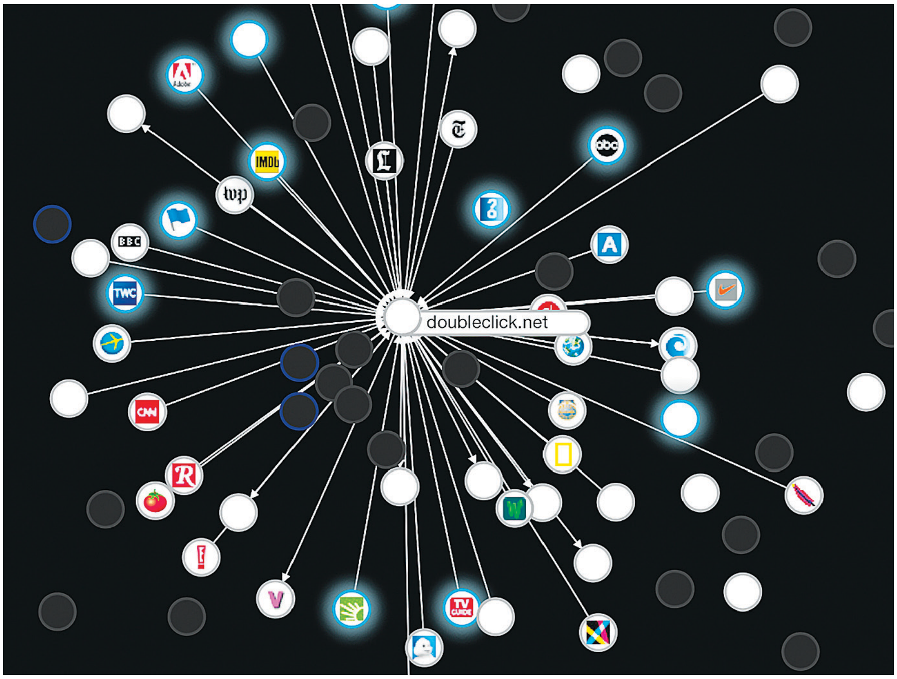
Society | DOI:10.1145/2240236.2240243

Samuel Greengard

Advertising Gets Personal

Online behavioral advertising and sophisticated data aggregation have changed the face of advertising and put privacy in the crosshairs.



Collusion, a Firefox add-on, lets a person see all the third-party entities tracking his or her movements across the Web.

SCREENSHOT COURTESY OF MOZILLA.ORG/COLLUSION

NOT LONG AGO, a man walked into the local Target store in Minneapolis and demanded to speak to the manager. He wanted to know why his then-high school daughter was receiving coupons and promotions for maternity clothing, cribs, and other items that would indicate she was pregnant. "Are you trying to encourage her to get pregnant?" he asked. The store manager examined the stack of coupons and promptly apologized. He said he did not have any idea why the girl had received the coupons. The man then left for home.

This could have been the end of the story. But after talking to his daughter the man discovered she was pregnant. Target had used sophisticated predictive analytics to determine that her previous buying patterns and behavior had indicated a high probability of expecting a baby. In fact, Target and other stores have become so good at gauging customers' buying patterns they now disguise customer-specific promotions by including coupons that are completely irrelevant to the recipient.

Welcome to the new world of advertising. As statisticians, software developers, and advertising experts mine and mix growing volumes of online and offline data and develop increasingly complex algorithms, they are building new and remarkably sophisticated advertising models designed to maximize results—and revenues. "Technology is enabling new—and in some cases hyper-local and personalized—forms of advertising," states John Nicholson, counsel for the law firm Pillsbury Winthrop Shaw Pittman. However, "there's a fine line between what's acceptable and what constitutes an invasion of privacy."

"Online behavioral advertising methods are advancing at an incredibly fast pace," says Lorrie Faith Cranor, an associate professor of computer science, engineering, and public policy at Carnegie Mellon University. "There are clearly advantages to receiving relevant ads, but the Internet, combined with today's data-collection technology, poses serious privacy concerns. Unfortunately, most consumers feel as though they have little control over what happens to their data and how it is used by advertisers."

By the Numbers

Over the years, advertisers have struggled to better understand the whims of the marketplace and target consumers more effectively. Identifying market niches and customer segments has been a daunting task and there has been no easy way to deliver relevant ads. The result? Most ads target broad demographic segments through television, radio, newspapers, magazines, kiosks, billboards, shopping carts, and other media. In many cases, advertisers simply hope for positive results and learn by trial and error.

However, the last few years have brought about a revolution in data mining particularly as online and conventional methods have allowed advertisers to assemble and reassemble data in new ways. A growing number of retailers, including Target, assign each customer a unique ID number or guest code. It is associated with a credit or debit card, and an individual's purchase history is stored for analysis. This is separate from a loyalty program, and the only way to avoid tracking is to pay with cash and avoid giving a phone number or any other personal data. But the process does not stop there. Increasingly, retailers and others plug in information from third-party sources that track the same individual. This might include the person's Web browsing patterns, credit history, what magazines they read, and even conversations they have had at social media sites.

The result is a fairly comprehensive picture of an individual's buying habits and consumption patterns. This profile—which could include anything from the type of tea or liquor a person likes to consume to medical conditions and sexual orientation—allows marketers to customize ads, but it also offers deep insights into life events and changes. For example, when a woman begins buying vitamin supplements, larger quantities of skin lotion, hand sanitizers, and a larger purse or bag there is an extremely high likelihood she is pregnant. In addition, analytics software has become so sophisticated it is possible to estimate the delivery window within a few weeks.

Of course, using data to predict life events could have far-reaching consequences, particularly if family, friends, or a prospective employer become aware of a sensitive lifestyle or medical issue, such as an affinity for nude beaches or a diagnosis of HIV. Worse, the data may contain errors and present an inaccurate picture that could lead to an employer refusing to hire the person or the loss of a job. As a result, advertisers are attempting to get smarter—some would say sneakier—in the way they deliver ads. Increasingly, they are including coupons and ads that are completely random or irrelevant in order to appear as though they are not spying over a person's shoulder.

Joseph Turian, president of consulting firm MetaOptimize, says that as organizations learn to use analytics and cultivate big data, insights that would have been unimaginable only a few years ago are moving into the mainstream. There are clear advantages for consumers—particularly those looking for discounts and deals—but advertisers need to avoid stepping over the line. "People like the idea of personalized searches and advertising," says Turian. "Many already provide data willingly for discounts through rewards programs. But they want to be in control of their destiny."

Cookies, Tweets, and Dollars

What makes the emerging field of data aggregation and analytics possible is a spate of online data-collection techniques that revolve around IP addresses, third-party cookies, and Web tools that track consumers as they click through Web sites and interact online. Internet service providers, Web sites, and advertising networks sell this data to other companies, including data aggregators. Google, meanwhile, collects data from searches and through keywords in Gmail and YouTube while Facebook has unlimited access to the mother lode of information and messages that appear on its site. Finally, Twitter recently sold its multibillion tweet archive to a U.K. firm that reportedly has more than 1,000 companies lined up for the data.

Today's data-collection system is largely based on an opt-out model that is nearly impossible to understand or manage, many privacy advocates contend. Consumers face the daunting task of trying to decipher lengthy and convoluted privacy policies that in some cases do not match actual practices, Cranor says. What is more, data collection firms often rely on loopholes and devious methods to circumvent cookie-blocking tools built into Web browsers and privacy tools such as Ghostery. In the end, users' attempts to control tracking and personal data often ends up resembling a game of Whac-A-Mole, Nicholson says.

In fact, half of all Internet users recall the ads they view but only 12% correctly remember the disclosure tag-lines attached to ads, Cranor reports. When she studied usage patterns she found that the majority of participants mistakenly believe that ads pop up if they click on disclosure icons and taglines. AdChoices, the tagline most commonly used by online advertisers (it discloses sites' advertising methods and allows consumers to click a button and opt out), was par-

Analytics software has become so sophisticated it is now possible to estimate a pregnant customer's delivery window within a few weeks.

Cloud Computing

New RaaS Pricing Model

The resources behind cloud computing services will soon be sold in increments of seconds, according to researchers from Technion-Israel Institute of Technology.

Providers of cloud computing have moved from renting servers on a monthly basis to renting virtual "server equivalents" for as little as an hour at a time. But even that is inefficient, say the researchers, who presented a paper, "The Resource-as-a-Service (RaaS) Cloud," at the recent USENIX Hotcloud '12 conference in Boston. Providers are moving toward pricing individual resources, such as memory, within a virtual machine, and changing prices in intervals of seconds, based on shifting demand. That trend from infrastructure-as-a-service to resource-as-a-service can save buyers money, earn more for providers, and make efficient use of hardware and energy.

"Clients don't need to buy things they don't need, hosts don't need to sell them things they don't need, and hosts can accommodate more clients on a server," says Orna Agmon Ben-Yehuda, a doctoral student and co-author of the paper.

Clients would use an "economic agent" that makes split-second decisions on how much to spend on which resources, and hosts would allocate resources based on how much a client was willing to pay. Cloud service providers' software could also incorporate economic agents to represent their own business interests.

Coauthor and doctoral student Muli Ben-Yehuda says the trend demands a lot of cloud computing researchers. Software, for instance, will have to adapt to use an ever-shifting set of resources, and workloads will need to be carefully balanced. The challenge, he says, is how to turn computing into a commodity. "How do you make computing something like electricity? It's there whenever you want it, you can have as much as you need, and the price is set by the market."

—Neil Savage

ticularly ineffective at communicating notice and choice. Nearly half of the participants who saw AdChoices believed it was intended to sell advertising space, while a mere 27% believed it was a means to stop tailored ads. "A majority of participants mistakenly believed that opting out would stop all online tracking, not just tailored ads," she notes.

Critics believe the inability to control what software and tracking mechanisms are placed on a person's computer is nothing less than a violation. Many Web sites contain a half-dozen to a dozen or more tracking tools or third-party cookies. It is akin to a company installing video cameras and microphones in a home and recording everything that occurs in the household. "When people find out what is really happening, the typical response is 'Are you kidding!'" says Marcella Wilson, an adjunct professor of computer science at the University of Maryland, Baltimore County.

Privacy Matters

In February, U.S. President Obama unveiled a Consumer Privacy Bill of Rights as part of a comprehensive blueprint to expand privacy protections while continuing to make the Internet a hub of innovation and economic growth. The measure attempts to force companies such as Google, Microsoft, and Yahoo! to stop monitoring when a person clicks a Do Not Track button on their Web browser. Do Not Track is intended to supersede a decade-old voluntary industry initiative called P3P, which has produced tepid results and proved unenforceable. It was designed to offer consumers some control over the type of data collected, how it could be used, and how long it could be stored.

Nicholson, who writes privacy policies for businesses, believes a fundamental problem lies in overly complex and incomprehensible privacy policies, as well as the way data is collected and used in the U.S. compared to Europe and other parts of the world. "The U.S. has treated personal information as more of a sales transaction and said that businesses can do what they want with it," he explains. "Europeans use more of a licensing model that focuses on the person owning their data and a business renting it for a specific use." In fact, the European Data Protection Directive and a newer Data Protection Regulation sets tight controls over how data can be collected, stored, and used. It also includes provisions for notifying consumers and obtaining their consent.

Nicholson says some privacy advocates now support a system modeled after Canada's Privacy by Design initiative, which aims to embed privacy protection into new technology and business processes by default. The underlying goal is for consumers to choose the data they make available to companies. With Privacy by Design, data aggregators would cull only the data they need and have permission to use, keep it for only as long as it is immediately valuable, and then purge the data, he explains.

Others have floated the idea of creating information exchanges that pay consumers for the use of their data. Essentially, an individual would manage his or her profile and decide who can purchase the data, what purposes they can use it for, and for how long. "The reality is that we're currently paying in a currency we don't understand because most people don't recognize the actual value of personal information," Nicholson explains.

Cranor says that, in the end, a balance must exist between today's rapidly advancing data-aggregation methods and increasingly elusive privacy. Although a strict opt-in model would almost certainly prove too unwieldy and annoying for consumers, "the entire process must be more transparent," says Cranor. "People must understand what is happening with their data and what choices they are actually making. Only then can we have a system that works well for everyone."

Further Reading

Leon, P.G., et al.
What do behavioral advertising disclosures communicate to users? Carnegie Mellon University CyLab, April 2, 2012.

Goldfarb, A. and Tucker, C.
Privacy regulation and online advertising, Social Science Research Network, August 4, 2010.

CNN
Joel Stein talks data mining and personal privacy on CNN American Morning, http://www.youtube.com/watch?v=rYiGT4EuE9w, March 10, 2011.

Tucker, C.E.
The economics of advertising and privacy, *International Journal of Industrial Organization 30*, 3, May 2012.

Turow, J.
***The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*, Yale University Press, New Haven, CT, 2012.**

Samuel Greengard is an author and journalist based in West Linn, OR.

Security

Target Cybercriminals, Urges Report

Governments should spend less money on defensive measures, such as antivirus software and firewalls, and more money on "hunting down cybercriminals and throwing them in jail," according to "Measuring the Cost of Cybercrime," a research paper by Ross Anderson, a professor of security engineering at the University of Cambridge, and a team of six researchers from Germany, the Netherlands, the U.K., and the U.S.

The paper, which the researchers believe is the first systematic report on the costs of cybercrime, was presented at the recent 11th Annual Workshop on the Economics of Information Security.

Developed at the request of the U.K. Ministry of Defense, the report provided estimates of the direct costs, indirect costs, and defense costs of different types of cybercrime in the U.K. and the world. It found that for Internet-based crimes like phishing, spam, online scams, hacking, and denial-of-service attacks, the costs of defense are many times higher than the actual losses. Anderson noted the example of a botnet that was responsible for a third of the world's spam in 2010. The botnet is estimated to have earned its owners about U.S. $2.7 million whereas the worldwide costs of spam prevention probably exceeded U.S. $1 billion.

—Jack Rosenberger