THE WALL STREET JOURNAL.

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers visit https://www.djreprints.com.

https://www.wsj.com/articles/SB10001424052748703940904575395073512989404

WHAT THEY KNOW

The Web's New Gold Mine: Your Secrets

A Journal investigation finds that one of the fastest-growing businesses on the Internet is the business of spying on consumers. First in a series.

By Julia Angwin

July 30, 2010

Hidden inside Ashley Hayes-Beaty's computer, a tiny file helps gather personal details about her, all to be put up for sale for a tenth of a penny.

The file consists of a single code— 4c812db292272995e5416a323e79bd37—that secretly identifies her as a 26-year-old female in Nashville, Tenn.

The code knows that her favorite movies include "The Princess Bride," "50 First Dates" and "10 Things I Hate About You." It knows she enjoys the "Sex and the City" series. It knows she browses entertainment news and likes to take quizzes.

"Well, I like to think I have some mystery left to me, but apparently not!" Ms. Hayes-Beaty said when told what that snippet of code reveals about her. "The profile is eerily correct."

Ms. Hayes-Beaty is being monitored by Lotame Solutions Inc., a New York company that uses sophisticated software called a "beacon" to capture what people are typing on a website—their comments on movies, say, or their interest in parenting and pregnancy. Lotame packages that data into profiles about individuals, without determining a person's name, and sells the profiles to companies seeking customers. Ms. Hayes-Beaty's tastes can be sold wholesale (a batch of movie lovers is $1 per thousand) or customized (26-year-old Southern fans of "50 First Dates").

"We can segment it all the way down to one person," says Eric Porres, Lotame's chief marketing officer.

One of the fastest-growing businesses on the Internet, a Wall Street Journal investigation has found, is the business of spying on Internet users.

Ashley Hayes-Beaty's taste in film is tracked by a New York firm—and offered for sale for a tenth of a cent. BRIAN MCCORD FOR THE WALL STREET JOURNAL

JOURNAL COMMUNITY »

The Journal conducted a comprehensive study that assesses and analyzes the broad array of cookies and other surveillance technology that companies are deploying on Internet users. It reveals that the tracking of consumers has grown both far more pervasive and far more intrusive than is realized by all but a handful of people in the vanguard of the industry.

• The study found that the nation's 50 top websites on average installed 64 pieces of tracking technology onto the computers of visitors, usually with no warning. A dozen sites each installed more than a hundred. The nonprofit Wikipedia installed none.

• Tracking technology is getting smarter and more intrusive. Monitoring used to be limited mainly to "cookie" files that record websites people visit. But the Journal found new tools that scan in real time what people are doing on a Web page, then instantly assess location, income, shopping interests and even medical conditions. Some tools surreptitiously re-spawn themselves even after users try to delete them.

• These profiles of individuals, constantly refreshed, are bought and sold on stock-market-like exchanges that have sprung up in the past 18 months.

The new technologies are transforming the Internet economy. Advertisers once primarily bought ads on specific Web pages—a car ad on a car site. Now, advertisers are paying a premium to follow people around the Internet, wherever they go, with highly specific marketing messages.

In between the Internet user and the advertiser, the Journal identified more than 100 middlemen—tracking companies, data brokers and advertising networks—competing to meet the growing demand for data on individual behavior and interests.

The data on Ms. Hayes-Beaty's film-watching habits, for instance, is being offered to advertisers on BlueKai Inc., one of the new data exchanges.

"It is a sea change in the way the industry works," says Omar Tawakol, CEO of BlueKai. "Advertisers want to buy access to people, not Web pages."

The Journal examined the 50 most popular U.S. websites, which account for about 40% of the Web pages viewed by Americans. (The Journal also tested its own site, WSJ.com.) It then analyzed the tracking files and programs these sites downloaded onto a test computer.

As a group, the top 50 sites placed 3,180 tracking files in total on the Journal's test computer. Nearly a third of these were innocuous, deployed to remember the password to a favorite site or tally most-popular articles.

But over two-thirds—2,224—were installed by 131 companies, many of which are in the business of tracking Web users to create rich databases of consumer profiles that can be sold.

The top venue for such technology, the Journal found, was IAC/InterActive Corp.'s Dictionary.com. A visit to the online dictionary site resulted in 234 files or programs being downloaded onto the Journal's test computer, 223 of which were from companies that track Web users.

The information that companies gather is anonymous, in the sense that Internet users are identified by a number assigned to their computer, not by a specific person's name. Lotame, for instance, says it doesn't know the name of users such as Ms. Hayes-Beaty—only their behavior and attributes, identified by code number. People who don't want to be tracked can remove themselves from Lotame's system.

And the industry says the data are used harmlessly. David Moore, chairman of 24/7 RealMedia Inc., an ad network owned by WPP PLC, says tracking gives Internet users better advertising.

"When an ad is targeted properly, it ceases to be an ad, it becomes important information," he says.

Tracking isn't new. But the technology is growing so powerful and ubiquitous that even some of America's biggest sites say they were unaware, until informed by the Journal, that they were installing intrusive files on visitors' computers.

DIG DEEPER

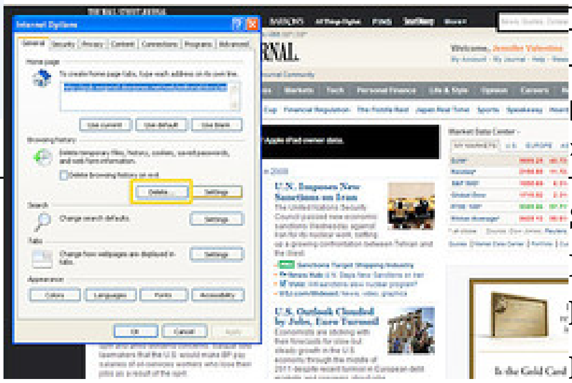
- Part 2 in Series: Microsoft Quashed Bid to Boost Web Privacy
- Part 3 in Series: On Web's Frontier, Anonymity in Name Only
- Part 4 in Series: Stalking by Cellphone
- Part 5 in Series: Google Agonizes Over Privacy
- Personal Details Exposed Via Biggest U.S. Websites
- The Journal's Methodology
- What They Know About You
- Digits: Your Questions on Digital Privacy
- Digits: Analyzing What You Have Typed
- Digits: Lawsuit Tackles Files That 'Re-Spawn' Cookies
- Full Coverage: wsj.com/WTK

GLOSSARY »

Key tracking terminology

HOW TO PROTECT YOURSELF »

Almost every major website you visit is tracking your online activity. Here's a step-by-step guide to fending off trackers.


Internet Options
Delete...
Settings
OK
Cancel
U.S. Imposes New Sanctions on Iran
U.S. Outlook Clouded by Jobs, Euro Turmoil

THE TRACKING ECOSYSTEM »

Surfing the Internet kickstarts a process that passes information about you and your interests to tracking companies and advertisers. See how it works.

The Journal found that Microsoft Corp.'s popular Web portal, MSN.com, planted a tracking file packed with data: It had a prediction of a surfer's age, ZIP Code and gender, plus a code containing estimates of income, marital status, presence of children and home ownership, according to the tracking company that created the file, Targus Information Corp.

Both Targus and Microsoft said they didn't know how the file got onto MSN.com, and added that the tool didn't contain "personally identifiable" information.

Tracking is done by tiny files and programs known as "cookies," "Flash cookies" and "beacons." They are placed on a computer when a user visits a website. U.S. courts have ruled that it is legal to deploy the simplest type, cookies, just as someone using a telephone might allow a friend to listen in on a conversation. Courts haven't ruled on the more complex trackers.

The most intrusive monitoring comes from what are known in the business as "third party" tracking files. They work like this: The first time a site is visited, it installs a tracking file, which assigns the computer a unique ID number. Later, when the user visits another site affiliated with the same tracking company, it can take note of where that user was before, and where he is now. This way, over time the company can build a robust profile.

One such ecosystem is Yahoo Inc.'s ad network, which collects fees by placing targeted advertisements on websites. Yahoo's network knows many things about recent high-school graduate Cate Reid. One is that she is a 13- to 18-year-old female interested in weight loss. Ms. Reid was able to determine this when a reporter showed her a little-known feature on Yahoo's website, the Ad Interest Manager, that displays some of the information Yahoo had collected about her.

Yahoo's take on Ms. Reid, who was 17 years old at the time, hit the mark: She was, in fact, worried that she may be 15 pounds too heavy for her 5-foot, 6-inch frame. She says she often does online research about weight loss.

"Every time I go on the Internet," she says, she sees weight-loss ads. "I'm self-conscious about my weight," says Ms. Reid, whose father asked that her hometown not be given. "I try not to think about it.... Then [the ads] make me start thinking about it."

Yahoo spokeswoman Amber Allman says Yahoo doesn't knowingly target weight-loss ads at people under 18, though it does target adults.

"It's likely this user received an untargeted ad," Ms. Allman says. It's also possible Ms. Reid saw ads targeted at her by other tracking companies.

Information about people's moment-to-moment thoughts and actions, as revealed by their online activity, can change hands quickly. Within seconds of visiting eBay . com or Expedia . com, information detailing a Web surfer's activity there is likely to be auctioned on the data exchange run by BlueKai, the Seattle startup.

Each day, BlueKai sells 50 million pieces of information like this about specific individuals' browsing habits, for as little as a tenth of a cent apiece. The auctions can happen instantly, as a website is visited.

Spokespeople for eBay Inc. and Expedia Inc. both say the profiles BlueKai sells are anonymous and the people aren't identified as visitors of their sites. BlueKai says its own website gives consumers an easy way to see what it monitors about them.

Tracking files get onto websites, and downloaded to a computer, in several ways. Often, companies simply pay sites to distribute their tracking files.

But tracking companies sometimes hide their files within free software offered to websites, or hide them within other tracking files or ads. When this happens, websites aren't always aware that they're installing the files on visitors' computers.

Often staffed by "quants," or math gurus with expertise in quantitative analysis, some tracking companies use probability algorithms to try to pair what they know about a person's online behavior with data from offline sources about household income, geography and education, among other things.

The goal is to make sophisticated assumptions in real time—plans for a summer vacation, the likelihood of repaying a loan—and sell those conclusions.

Some financial companies are starting to use this formula to show entirely different pages to visitors, based on assumptions about their income and education levels.

Life-insurance site AccuquoteLife.com, a unit of Byron Udell & Associates Inc., last month tested a system showing visitors it determined to be suburban, college-educated baby-boomers a default policy of $2 million to $3 million, says Accuquote executive Sean Cheyney. A rural, working-class senior citizen might see a default policy for $250,000, he says.

"We're driving people down different lanes of the highway," Mr. Cheyney says.

Consumer tracking is the foundation of an online advertising economy that racked up $23 billion in ad spending last year. Tracking activity is exploding. Researchers at AT&T Labs and Worcester Polytechnic Institute last fall found tracking technology on 80% of 1,000 popular sites, up from 40% of those sites in 2005.

The Journal found tracking files that collect sensitive health and financial data. On Encyclopaedia Britannica Inc.'s dictionary website Merriam-Webster.com, one tracking file from Healthline Networks Inc., an ad network, scans the page a user is viewing and targets ads related to what it sees there. So, for example, a person looking up depression-related words could see Healthline ads for depression treatments on that page—and on subsequent pages viewed on other sites.

Healthline says it doesn't let advertisers track users around the Internet who have viewed sensitive topics such as HIV/AIDS, sexually transmitted diseases, eating disorders and impotence. The company does let advertisers track people with bipolar disorder, overactive bladder and anxiety, according to its marketing materials.

Targeted ads can get personal. Last year, Julia Preston, a 32-year-old education-software designer in Austin, Texas, researched uterine disorders online. Soon after, she started noticing fertility ads on sites she visited. She now knows she doesn't have a disorder, but still gets the ads.

It's "unnerving," she says.

Tracking became possible in 1994 when the tiny text files called cookies were introduced in an early browser, Netscape Navigator. Their purpose was user convenience: remembering contents of Web shopping carts.

Back then, online advertising barely existed. The first banner ad appeared the same year. When online ads got rolling during the dot-com boom of the late 1990s, advertisers were buying ads based on proximity to content—shoe ads on fashion sites.

The dot-com bust triggered a power shift in online advertising, away from websites and toward advertisers. Advertisers began paying for ads only if someone clicked on them. Sites and ad networks began using cookies aggressively in hopes of showing ads to people most likely to click on them, thus getting paid.

Targeted ads command a premium. Last year, the average cost of a targeted ad was $4.12 per thousand viewers, compared with $1.98 per thousand viewers for an untargeted ad, according to an ad-industry-sponsored study in March.

The Journal examined three kinds of tracking technology—basic cookies as well as more powerful "Flash cookies" and bits of software code called "beacons."

More than half of the sites examined by the Journal installed 23 or more "third party" cookies. Dictionary.com installed the most, placing 159 third-party cookies.

Cookies are typically used by tracking companies to build lists of pages visited from a specific computer. A newer type of technology, beacons, can watch even more activity.

Beacons, also known as "Web bugs" and "pixels," are small pieces of software that run on a Web page. They can track what a user is doing on the page, including what is being typed or where the mouse is moving.

The majority of sites examined by the Journal placed at least seven beacons from outside companies. Dictionary.com had the most, 41, including several from companies that track health conditions and one that says it can target consumers by dozens of factors, including zip code and race.

Dictionary.com President Shravan Goli attributed the presence of so many tracking tools to the fact that the site was working with a large number of ad networks, each of which places its own cookies and beacons. After the Journal contacted the company, it cut the number of networks it uses and beefed up its privacy policy to more fully disclose its practices.

The widespread use of Adobe Systems Inc.'s Flash software to play videos online offers another opportunity to track people. Flash cookies originally were meant to remember users' preferences, such as volume settings for online videos.

But Flash cookies can also be used by data collectors to re-install regular cookies that a user has deleted. This can circumvent a user's attempt to avoid being tracked online. Adobe condemns the practice.

Most sites examined by the Journal installed no Flash cookies. Comcast . net installed 55.

That finding surprised the company, which said it was unaware of them. Comcast Corp. subsequently determined that it had used a piece of free software from a company called Clearspring Technologies Inc. to display a slideshow of celebrity photos on Comcast.net. The Flash cookies were installed on Comcast's site by that slideshow, according to Comcast.

Clearspring, based in McLean, Va., says the 55 Flash cookies were a mistake. The company says it no longer uses Flash cookies for tracking.

CEO Hooman Radfar says Clearspring provides software and services to websites at no charge. In exchange, Clearspring collects data on consumers. It plans eventually to sell the data it collects to advertisers, he says, so that site users can be shown "ads that don't suck." Comcast's data won't be used, Clearspring says.

Wittingly or not, people pay a price in reduced privacy for the information and services they receive online. Dictionary.com, the site with the most tracking files, is a case study.

The site's annual revenue, about $9 million in 2009 according to an SEC filing, means the site is too small to support an extensive ad-sales team. So it needs to rely on the national ad-placing networks, whose business model is built on tracking.

> Think about how these technologies and the associated analytics can be used in other industries and social settings (e.g. education) for real beneficial impacts. This is nothing new for the web, the now that it has matured, it can be a positive game-changer.
>
> *—Mitchell Weisberg*

Dictionary.com executives say the trade-off is fair for their users, who get free access to its dictionary and thesaurus service.

"Whether it's one or 10 cookies, it doesn't have any impact on the customer experience, and we disclose we do it," says Dictionary.com spokesman Nicholas Graham. "So what's the beef?"

The problem, say some industry veterans, is that so much consumer data is now up for sale, and there are no legal limits on how that data can be used.

Until recently, targeting consumers by health or financial status was considered off-limits by many large Internet ad companies. Now, some aim to take targeting to a new level by tapping online social networks.

Media6Degrees Inc., whose technology was found on three sites by the Journal, is pitching banks to use its data to size up consumers based on their social connections. The idea is that the creditworthy tend to hang out with the creditworthy, and deadbeats with deadbeats.

"There are applications of this technology that can be very powerful," says Tom Phillips, CEO of Media6Degrees. "Who knows how far we'd take it?"

—Emily Steel, Jennifer Valentino-DeVries and Tom McGinty contributed to this report.

Write to Julia Angwin at julia.angwin@wsj.com

Follow @whattheyknow on Twitter

- Microsoft Quashed Bid to Boost Web Privacy
- Top Sites Feed Personal Data

How to Avoid Prying Eyes

What They Know About You

Analyzing What You Have Typed

Decoding the Trackers: A Glossary

The Journal's Methodology

Copyright © 2019 Dow Jones & Company, Inc. All Rights Reserved

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to your colleagues, clients or customers visit https://www.djreprints.com.