THE WALL STREET JOURNAL.
WSJ.com

What They Know

The Business of Tracking You on the Internet

A Wall Street Journal Investigation

THE WALL STREET JOURNAL.
WSJ.com

Table of Contents

THE WALL STREET JOURNAL.
WSJ.com

Introduction

What do they know about me?

Who hasn't wondered that after receiving a spookily correct mortgage offer for the exact balance of your debt? Or when an ad appears on your computer screen for something you were just thinking about but hadn't typed yet?

Our modern life is filled with these minor but eerie intrusions – whether they arrive in snail mail, the e-mail inbox, or pop up as we browse the Web. Marketers are indeed watching us and compiling dossiers about us all the time. But until recently, there wasn't an easy way to find out what they knew about us.

Until the Internet came along, the information that companies had about us was stored in dusty files and used primarily by the folks filling our mailboxes with junk mail. Gathering information about people was expensive and difficult: data brokers had to gather people's real estate records, motor vehicle records and keep up with Americans' propensity to move every few years.

The Internet made monitoring people much easier. Suddenly companies could embed a tiny piece of code in a website and see anything that a person was doing on a web page. As people spent more time online, more of their life could be monitored.

By 2010, online tracking had become a fundamental part of the $23 billion online advertising economy. Hundreds of companies popped up to offer new ways to track users online. Trading floors emerged to allow our digital records to be bought and sold in an instant. And the beauty of these digital files was that occasionally they could be decoded – revealing a dossier for the first time.

Most people surfing the Web had no idea of the scope and intrusiveness of this new industry that was watching their every move.

So The Wall Street Journal sought to decode these new tracking technologies, allowing readers for the first time to glimpse behind the curtain of the personal data-gathering industry.

It wasn't an easy task. The Journal reporters learned how to "sniff on the wire" – or decode computer talk – to identify tracking tools on the 100 of the most popular kids and adult websites. The Journal hired technologists to set up a

mobile lab in Denver to test "apps" in a secure environment – disabling the phones' cellular service and forcing traffic through Wi-Fi where it could be collected and analyzed in a groundbreaking study. Journal reporters cracked computer code to reveal a significant Facebook privacy breach, and how Capital One was using tracking data to estimate the incomes and lifestyles of visitors to its website.

The series launched on July 31, 2010, with an online database of all the tracking tools the Journal had found in its visits to the top 50 U.S. websites. Its findings included: 234 tracking devices on Dictionary.com alone; companies estimating income and diseases of users; and a company tracking people's favorite movies as they typed them into a website.

The Journal's reporting shocked even the technology elite, many of whom hadn't realized how sophisticated the tracking industry had become. "It's pretty freaking amazing — and amazingly freaky," Doc Searls, a fellow at Harvard University's Berkman Center for Internet and Society, wrote on his blog the day the series launched. "The tide has turned today."

The tide continued to turn throughout the year, as the Journal revealed more disturbing facts about the commercial data-gathering industry. It caught red-handed a company breaking Facebook's rules to obtain user names for sale. It nabbed Nielsen Co. breaking into a medical website to "scrape" patient data and sell it.

It revealed companies using undetectable tracking techniques, such as "deep packet inspection" and "device fingerprinting." It found that advertisers had influenced Microsoft's decision to remove privacy features from its Web browser.

It found that children's websites were more heavily monitored than adults'. It found that iPhone and Android "apps" were secretly sending out data about users to tracking companies. And the Journal found that tracking data was being used by life insurers and credit card issuers to help make financial decisions about customers.

In short, the Journal unveiled a massive surveillance industry using sophisticated tools to secretly monitor users' behavior – and to use that information to make important decisions about people's lives.

The impact of the series was profound. Driven by swelling public concern about tracking, the Obama administration reversed the government's decade-long hands-off approach to Internet privacy regulation and called for an Internet "privacy bill of rights." The Federal Trade Commission, which had previously supported the industry's self-regulatory efforts, declared in December that self-regulation had failed and called for a do-not-track tool to be installed in Web browsing software.

Companies also began changing their privacy practices in response to the Journal's reporting. Facebook banned the data collection company RapLeaf Inc.

from its website after the Journal revealed that RapLeaf was taking user information and transmitting it to tracking companies. After the Journal's article, Nielsen said it would no longer create fake usernames and passwords to log into private message boards to scrape data.

In December, Microsoft Corp. reversed its decision to remove privacy tools from its Web browser. It will add a powerful privacy feature similar to the one it dropped from an earlier version back into Internet Explorer 9 when it launches in 2011. Mozilla Corp. soon followed by announcing it would add a do-not-track tool to the Firefox Web browser. And Google said it would improve an anti-tracking tool it offered.

And the debate about tracking is only beginning. Online tracking tools are getting ever more sophisticated. Ultimately, we will need to decide how much surveillance we as a society are willing to accept.

JULIA ANGWIN
Senior technology editor, WSJ.com

THE WALL STREET JOURNAL.
WSJ.com

Contributors

Reporters

Julia Angwin, Geoffrey Fowler, Yukari Iwatani Kane, Mark Maremont, Justin Scheck, Leslie Scism, Paul Sonne, Steve Stecklow, Emily Steel, Scott Thurm, Jennifer Valentino-DeVries, Jessica Vascellaro, Nick Wingfield

Editors

Jesse Pesta, Julia Angwin, Scott Thurm, Steve Yoder, Mitch Pacelle

Research and data analysis

Tom McGinty, Julia Angwin, Courtney Banks, Marisa Taylor, Scott Thurm, Jennifer Valentino-DeVries

Print and interactive graphics

Paul Antonson, Andrew Garcia-Phillips, Mei Lan Ho-Walker, Jovi Juan, Jonathan Keegan, Susan McGregor, Andrew Robinson, Sarah Slobin, Kurt Wilberding

Technology consultants

David Campbell, Ashkan Soltani

THE WALL STREET JOURNAL.
WSJ.com

The Web's New Gold Mine: Your Secrets

BY JULIA ANGWIN

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.

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.

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., in June 2010 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 in 2009. 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 Encyclopedia 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. In 2009, 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. In 2009, 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 2010.

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.

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.

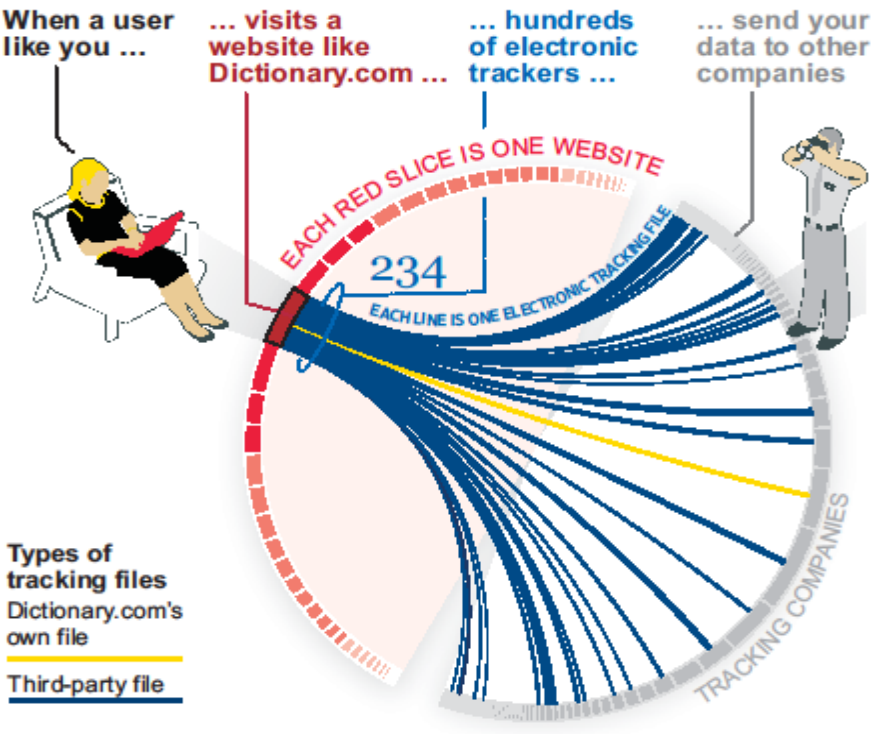
"There are applications of this technology that can be very powerful," says Tom Phillips, CEO of Media6Degrees. "Who knows how far we'd take it?"

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

Published July 30, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Explore the Data



A WSJ study isolated this universe of 3,180 tracking files on the Web's biggest sites. Among the Web's 50 most-popular U.S. sites, **Dictionary.com** hosted the most trackers and was least protective of its users' information.

The 'exposure index' was calculated based on a scoring of those tracking files and the total number of files on that site.

How much each website exposes user data ...

MOST EXPOSURE ⟷ LEAST EXPOSURE

EXPLORE ALL THE SITES IN THE DATABASE:

http://blogs.wsj.com/wtk/

THE WALL STREET JOURNAL.
WSJ.com

Sites Feed Personal Details To New Tracking Industry

BY JULIA ANGWIN and TOM MCGINTY

The largest U.S. websites are installing new and intrusive consumer-tracking technologies on the computers of people visiting their sites—in some cases, more than 100 tracking tools at a time—a Wall Street Journal investigation has found.

The tracking files represent the leading edge of a lightly regulated, emerging industry of data-gatherers who are in effect establishing a new business model for the Internet: one based on intensive surveillance of people to sell data about, and predictions of, their interests and activities, in real time.

The Journal's study shows the extent to which Web users are in effect exchanging personal data for the broad access to information and services that is a defining feature of the Internet.

In an effort to quantify the reach and sophistication of the tracking industry, the Journal examined the 50 most popular websites in the U.S. to measure the quantity and capabilities of the "cookies," "beacons" and other trackers installed on a visitor's computer by each site. Together, the 50 sites account for roughly 40% of U.S. page-views.

The 50 sites installed a total of 3,180 tracking files on a test computer used to conduct the study. Only one site, the encyclopedia Wikipedia.org, installed none. Twelve sites, including IAC/InterActive Corp.'s Dictionary.com, Comcast Corp.'s Comcast.net and Microsoft Corp.'s MSN.com, installed more than 100 tracking tools apiece in the course of the Journal's test.

The Journal also surveyed its own site, WSJ.com, which doesn't rank among the top 50 by visitors. WSJ.com installed 60 tracking files, slightly below the 64 average for the top 50 sites.

Some two-thirds of the tracking tools installed—2,224—came from 131 companies that, for the most part, are in the business of following Internet users to create rich databases of consumer profiles that can be sold. The companies that placed the most such tools were Google Inc., Microsoft Corp and Quantcast Corp., all of which are in the business of targeting ads at people online.

Google, Microsoft and Quantcast all said they don't track individuals by name and offer Internet users a way to remove themselves from their tracking networks. Comcast, MSN and Dictionary.com said they disclose tracking practices in their privacy policies, and said their visitors aren't identified by name.

The state of the art is growing increasingly intrusive, the Journal found. Some tracking files can record a person's keystrokes online and then transmit the text to a data-gathering company that analyzes it for content, tone and clues to a person's social connections. Other tracking files can re-spawn trackers that a person may have deleted.

To measure the sensitivity of the data gathered by tracking companies, the Journal created an "exposure index" for the top 50 sites. Dictionary.com ranked highest in exposing users to potentially aggressive surveillance: It installed 168 tracking tools that didn't let users decline to be tracked, and 121 tools that, according to their privacy statements, don't rule out collecting financial or health data. Dictionary.com attributed the number of tools to its use of many different ad networks, each of which puts tools on its site.

Some of the tracking files identified by the Journal were so detailed that they verged on being anonymous in name only. They enabled data-gathering companies to build personal profiles that could include age, gender, race, zip code, income, marital status and health concerns, along with recent purchases and favorite TV shows and movies.

The ad industry says tracking doesn't violate anyone's privacy because the data sold doesn't identify people by name, and the tracking activity is disclosed in privacy policies. And while many companies are involved in collecting, analyzing and selling the data, they provide a useful service by raising the chance Internet users see ads and information relevant to them personally.

"We are delivering free content to consumers," says Mike Zaneis, vice president of public policy for the Interactive Advertising Bureau, a trade group of advertisers and publishers. "Sometimes it means that we get involved in a very complex ecosystem with lots of third parties."

The growing use and power of tracking technology have begun to raise regulatory concerns. Congress is considering laws to limit tracking. The Federal Trade Commission is developing privacy guidelines for the industry.

If "you were in the Gap, and the sales associate said to you, 'OK, from now on, since you shopped here today, we are going to follow you around the mall and view your consumer transactions,' no person would ever agree to that," Sen. George LeMieux, R-Florida, said this week in a Senate hearing on Internet privacy.

Published July 30, 2010.

THE WALL STREET JOURNAL.
WSJ.com

How to Avoid the Prying Eyes

BY JENNIFER VALENTINO-DEVRIES

Visitors to almost every major website are tracked online, a Journal investigation has found. But there are ways to limit the snooping.

Web browsing activity is tracked by use of "cookies," "beacons" and "Flash cookies," small computer files or software programs installed on a user's computer by the Web pages that are visited. Some are useful. But a subset ("third party" cookies and beacons) are used by companies to track users from site to site and build a database of their online activities.

Simple steps

Major browsers including Microsoft Corp.'s Internet Explorer, Mozilla Foundation's Firefox, Google Inc.'s Chrome and Apple Inc.'s Safari, have privacy features. To have the most privacy options, upgrade to the latest version of the browser you use.

Check and delete cookies: All popular browsers let users view and delete cookies installed on their computer. Methods vary by browser.

For instance on Internet Explorer 8 (the most widely used browser), go to the "Tools" menu, pull down to "Internet Options" and under the "General" tab there are options for deleting some or all cookies. There might be hundreds, so deleting all might be easiest. But the next time you visit a favorite site, you may need to retype passwords or other login data previously stored automatically by one of those cookies.

For guides for all major browsers, go to WSJ.com/WTK.

Adjust Browser Settings: Once you've deleted cookies, you can limit the installation of new ones. Major browsers let you accept some cookies and block others. To maintain logins and settings for sites you visit regularly, but limit tracking, block "third-party" cookies. Safari automatically does this; other browsers must be set manually.

There are downsides to blocking all cookies. If you frequent sites that require logins, you will have to log in each time you visit.

Internet Explorer lets you set rules for blocking cookies based on the policies of the cookie-placer. One option blocks cookies that don't include a

privacy policy; another blocks cookies that can save your contact information without your approval. The control is under "Tools/Internet Options/Privacy."

No major browsers let you track or block beacons without installing extra software known as "plug-ins," as described under advanced steps.

Turn On "Private" Browsing: All major browsers offer a "private browsing" mode to limit cookies. Chrome calls it "Incognito." Internet Explorer calls it "InPrivate Browsing," but this option is available only in the latest version, IE8.

Private browsing doesn't block cookies. It deletes cookies each time you close the browser or turn off private browsing, effectively hiding your history.

Private browsing isn't selective. It deletes all cookies, whether useful or not. So you might want to use private browsing selectively, such as when looking at health-related information.

Monitor "Flash Cookies": Another kind of cookie uses Adobe Systems Inc.'s popular Flash program to save information on your computer. Flash is the most common way to show video online. As with regular cookies, Flash cookies can be useful for remembering preferences, such as volume settings for videos. But marketers also can use Flash cookies to track what you do online.

To identify the Flash cookies on your computer and adjust settings, go to: http://www.macromedia.com/support/documentation/en/flashplayer/help/settings_manager.html. You can delete Flash cookies stored on your computer and specify whether you want to accept future third-party Flash cookies.

The downside of blocking third-party Flash cookies: Some sites won't let you watch videos or other content.

Advanced Steps

Install Privacy "Plug-ins": Small programs called "add-ons" or "plug-ins" can help maintain privacy. Some let you monitor trackers that can't be seen through the browser; others allow you to delete cookies on a regular schedule.

Not all browsers can use all plug-ins. And some plug-ins can be tricky to set up. With those caveats, some plug-ins may be worth a look:

Abine: Developed by a Cambridge, Mass., start-up of the same name, it attempts to control several types of trackers. Once installed, the program will warn you when a site is placing cookies or Flash cookies on your machine. You can also see and block a third type of tracker called a Web "beacon" (sometimes called a "bug"). This is an invisible object embedded in a page that can interact with cookies. It's available only in "test" versions, so this is only for people who don't mind experimenting a bit with software. For Firefox, go to http://addons.mozilla.org/en-US/firefox/addon/11073/. For Internet Explorer, users need to request an invitation at getabine.com.

Better Privacy: This plug-in offers control over Flash cookies. It doesn't block them, but lets you set rules for deleting them—a distinction that can be

helpful if you frequent sites that require you to use third-party Flash cookies to see their content. Better Privacy (available only for Firefox) is at addons.mozilla.org/en-US/firefox/addon/6623/.

Ghostery: Available at ghostery.com ghostery.com, it helps control beacons. It alerts you when there's a beacon on a page you're viewing, tells you who placed it and details the company's privacy policy. With Internet Explorer or Firefox, you can then block the beacon from capturing information on your computer. That feature isn't available for Chrome.

Controlling Ads

Users troubled by targeted advertising can block or limit the ads being shown. Note: These tools don't necessarily restrict tracking. Some ad networks may still collect data on your browsing behavior and share it with others, even if you instruct them not to show you targeted ads.

The Network Advertising Initiative, an industry group of marketing companies, lets computer users opt out of targeted ads from about 50 ad networks at networkadvertising.org.

If you opt out, you won't be shown ads tied to your browsing behavior from the member networks. But you'll still see ads, which may be placed based on criteria such as your location.

PrivacyChoice LLC, an independent group, maintains a Web site (privacychoice.org/choose) that covers 152 ad networks. You can opt out of most by clicking a button there. For some, you'll need to download a plug-in, but it works only with Firefox.

Ironically, these opt-out systems work by installing a cookie on your computer. That cookie tells ad networks to stop sending targeted ads to your computer. Because these systems rely on a cookie to work, you'll need to opt out all over again any time you delete cookies from your machine.

Published July 30, 2010.

For interactive graphics related to this story, click this link.

THE WALL STREET JOURNAL.
WSJ.com

What They Know About You

BY JENNIFER VALENTINO-DEVRIES

A few online marketers will show you what they know about you—or think they know.

Google Inc., Microsoft Corp., Yahoo Inc. and others have created "preference managers" that let you see, and change, the interests they've assigned to you based on your browsing behavior. The companies acted partly in response to concerns about the privacy of the people they're tracking.

Some, but not all, of the preference managers let you halt tracking by that company. But none will block all tracking, or prevent you from seeing ads. Some require you to register, and one lets you pick a charity to which you can donate some of the money made by selling your data.

The companies gather this information by tracking your Web-surfing activity through small computer files or software programs installed on your computer by the websites you visit. Over time, this information says a lot about your interests. Companies then use the information to make other guesses and predictions about you, ranging from gender and age to marital status and creditworthiness.

Some of their guesses can be wrong. Your profile might also be inaccurate if you block or delete cookies, since that prevents companies from following your Web surfing to assess your interests.

Here's where you can go to see what the marketers think they know about you. Some let you correct or delete information already in their database about you, or add more information.

- Google was the most prevalent tracker in The Wall Street Journal's survey of popular websites, with tracking code appearing on 49 of the 50 sites tested. Google's Ads Preferences Manager, at http://www.google.com/ads/preferences/, lets you see the name of the tracking file, or "cookie," it associates with your Web browser and the interests it links to that cookie.

  Google's tool lets you remove some interests and add others, choosing from a list ranging from "poetry" to "neuroscience" to "polar regions." You can also disable Google's cookie so it no longer tracks you.
- The Microsoft Advertising tool

- The Microsoft ad tool, https://choice.live.com/UserPreferences, offers 1,295 interests to select, which will influence the ads you see. To change the selections, you must have an account with Hotmail or Windows Live ID. You can opt out of targeted ads from Microsoft without an account.
- Yahoo Inc.'s Ad Interest Manager, which is in testing, shows you the categories in which it thinks you have interests and lets you turn categories off; you can't add interests. Yahoo will also show you what it knows about your computer, including the operating system you are using and your browser, as well as the categories you search and the pages you visit on Yahoo.
  You can opt out of interest-based advertising. The tool is at http://privacy.yahoo.com/aim. Some less well-known companies that serve as middlemen for collecting and disseminating information about Internet users also offer preference managers.
- BlueKai Inc., a "data exchange" where marketers can buy and sell information about people's browsing habits, has a consumer-preferences registry at http://tags.bluekai.com/registry. The registry can include information about where you live, presumed interests and other guesses based on your browsing behavior, such as the kind of job you have and your approximate net worth.
  You can remove categories associated with your profile. You can also select a charity that will receive a donation based on advertising sales tied to your data, BlueKai says. You can stop BlueKai from collecting or trading information about you.
- EXelate Media, a marketplace for behavioral-targeting data, provides a preference manager at http://www.exelate.com/new/consumers-optoutpreferencemanager.html. The tool will show categories, such as "Asian Community" or "Casual Gaming," in which eXelate thinks you're interested. You can select the information that you want associated with you and remove what you don't want, or you can opt out of eXelate's program.
- Lotame Solutions Inc., a marketing technology company, lets you remove and add interests. You can indicate your gender and age, or opt out of receiving targeted ads. Its tool is available at http://www.lotame.com/preferences.html.

Published July 31, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Microsoft Quashed Effort To Boost Online Privacy

BY NICK WINGFIELD

The online habits of most people who use the world's dominant Web browser are an open book to advertisers. That wasn't the plan at first.

In early 2008, Microsoft Corp.'s product planners for the Internet Explorer 8.0 browser intended to give users a simple, effective way to avoid being tracked online. They wanted to design the software to automatically thwart common tracking tools, unless a user deliberately switched to settings affording less privacy.

That triggered heated debate inside Microsoft. As the leading maker of Web browsers, the gateway software to the Internet, Microsoft must balance conflicting interests: helping people surf the Web with its browser to keep their mouse clicks private, and helping advertisers who want to see those clicks.

In the end, the product planners lost a key part of the debate. The winners: executives who argued that giving automatic privacy to consumers would make it tougher for Microsoft to profit from selling online ads. Microsoft built its browser so that users must deliberately turn on privacy settings every time they start up the software.

Microsoft's original privacy plans for the new Explorer were "industry-leading" and technically superior to privacy features in earlier browsers, says Simon Davies, a privacy-rights advocate in the U.K. whom Microsoft consulted while forming its browser privacy plans. Most users of the final product aren't even aware its privacy settings are available, he says. "That's where the disappointment lies."

Microsoft General Counsel Brad Smith says that in developing the new browsers, the company tried to "synthesize" both points of view about privacy "in a way that advanced both the privacy interests of consumers and the critical role advertising plays in content."

Microsoft's decision reveals the economic forces driving the spread of online tracking of individuals. A Wall Street Journal investigation of the practice showed tracking to be pervasive and ever-more intrusive: The 50 most-popular U.S. websites, including four run by Microsoft, installed an average of 64 pieces of tracking technology each onto a test computer.

As online advertising grows more sophisticated, companies playing prominent roles in consumers' online experiences have discovered they have access to a valuable trove of information. In addition to Microsoft, such companies include search-engine giant Google Inc., iPhone maker Apple Inc., and Adobe Systems Inc., whose Flash software makes much of the Internet's video, gaming and animation possible. These companies now have a big say in how much information can be collected about individual users.

Many also have big stakes in online advertising. Microsoft bought aQuantive, a Web-ad firm, in 2007 for more than $6 billion, to build a business selling ads online. Google, already a giant in online marketing, in September 2008 launched a Web browser, Chrome, that gives it new insight into Internet users' habits. Apple has launched an ad network, iAds, for its iPhone and iPad. And Adobe in 2009 paid $1.8 billion to buy Omniture, which measures the effectiveness of online ads.

Executives in Microsoft's new ad business were upset when the designers of Internet Explorer hatched the plan to block tracking activity, say people involved in the debate. At a meeting in the spring of 2008, Brian McAndrews, a Microsoft senior vice president who had been chief executive of aQuantive before Microsoft acquired it, complained to the browser planners. Their privacy plan, he argued, would disrupt the selling of Web ads by Microsoft and other companies, these people say.

Mr. McAndrews was taken aback that Explorer planners seemed unwilling to accept input from advertising executives, given that Microsoft had spent $6 billion on a Web-ad firm, according two people who participated in the meeting.

Mr. Smith, the general counsel, says Microsoft weighed both sides of the argument in its debate. He says the company was concerned about the effect strict privacy features might have on free sites supported by advertising, including newspaper sites. Such sites, including WSJ.com, use information derived from tracking to sell targeted ads, an important revenue source.

Web browsers like Internet Explorer can play an important role in protecting privacy because the software sits between consumers and the array of technologies used to track them online. The best-known of those technologies are browser "cookies," small files stored on users' computers that act as identification tags for them when they visit websites.

Some cookies, such as those installed when a user asks a favorite website to remember his password, don't do tracking.

Others are installed on computers by companies that provide advertising services to the websites a user visits. These "third-party" cookies can be designed to track a user's online activities over time, building a database of personal interests and other details.

The Journal's examination of the top 50 most popular U.S. websites showed that Microsoft placed third-party tracking devices on 27 of the top 46 sites that it doesn't itself own.

All the latest Web browsers, including Internet Explorer, let consumers turn on a feature that prevents third-party browser cookies from being installed on their computers. But those settings aren't always easy to find. Only one major browser, Apple's Safari, is preset to block all third-party cookies, in the interest of user privacy.

"Only browser developers have the resources and large user bases necessary to create a privacy-friendly version of the Web," says Peter Eckersley, staff technologist with the Electronic Frontier Foundation, a digital-rights advocacy group.

Because Internet Explorer is used by so many people—nearly 60% of all Web users—the 2008 decision by planners of the new version to make it easy for users to block tracking could have had a big effect on the marketplace.

At the time, the practice of tailoring ads to consumers based on their browsing habits was taking off. Google was in the process of buying DoubleClick Inc., a leader in the placing and tracking of online ads, for $3.1 billion. A coalition of privacy groups was petitioning the Federal Trade Commission to develop stricter policies for preventing advertisers from tracking Web-browsing habits. Companies with stakes in Internet advertising were feeling heat to try to stave off government regulation by voluntarily protecting consumer privacy.

Microsoft also was trying to stem the erosion of its browser market share. Internet Explorer, which once had more than 95% of the market, hadn't kept up with competitors. Firefox, a Web browser overseen by the nonprofit Mozilla Foundation, picked up more than 18% of the market by May 2008, helping knock Explorer to 76%, according to NetApplications.com, which tracks browser use.

The browser planners at Microsoft believed aggressive new privacy features could help differentiate the new Internet Explorer from rivals, according to several current and former Microsoft executives.

The planners, led by Microsoft veteran Dean Hachamovitch, came up with a concept for preventing consumer tracking. A new feature would monitor where each piece of content on a visited Web page was originating on the Internet—every picture, video or chunk of text. The feature would pay special attention to content from "third party" Internet addresses—addresses different from the one a user sees in the address bar at the top of the browser.

Some of that third-party content could be innocuous things like YouTube video clips displayed on the Web page, which viewers presumably wouldn't want to block. Other items might be tracking tools such as Web "beacons," snippets of code embedded in the page that can monitor the clicks of visitors, or even record their keystrokes. Users might want such tracking tools blocked automatically.

The Internet Explorer planners proposed a feature that would block any third-party content that turned up on more than 10 visited websites, figuring that anything so pervasive was likely to be a tracking tool. This, they believed, was a more comprehensive approach to privacy than simply turning off browser cookies, one that would thwart other tracking methods.

The group also planned to design the Internet Explorer set-up process so that it guaranteed the privacy feature would be used by most people.

When he heard of the ideas, Mr. McAndrews, the executive involved with Microsoft's Internet advertising business, was angry, according to several people familiar with the matter. Mr. McAndrews feared the Explorer group's privacy plans would dramatically reduce the effectiveness of online advertising by curbing the data that could be collected about consumers.

He heard about the proposal through back channels rather than directly from the browser planners, these people say, which surprised him given its implications. Some people who worked in the browser group acknowledge that they should have been more upfront about their intentions. Mr. McAndrews later left the company.

"We were worried it was going to cause a stampede" away from tracking technologies, says an executive who worked with Mr. McAndrews. "It was an act with the potential to reverberate across the industry."

The browser group and its manager, Mr. Hachamovitch, tried to hold their ground. They were reluctant to let advertising executives interfere with the new Explorer design, according to people involved in the debate. Microsoft said that Mr. Hachamovitch and other members of the planning group wouldn't comment on the matter.

The debate widened after executives from Microsoft's advertising team informed outside advertising and online-publishing groups of Microsoft's privacy plans for Explorer. Microsoft Chief Executive Steve Ballmer assigned two senior executives, chief research and strategy officer Craig Mundie and the general counsel, Mr. Smith, to help referee the debate, according to Peter Cullen, Microsoft's chief privacy strategist.

The two men convened a four-hour meeting in Mr. Mundie's conference room in late spring 2008 to allow outside organizations to voice their concerns, including the Interactive Advertising Bureau, the Online Publishers' Association and the American Association of Advertising Agencies.

One of the attendees, Interactive Advertising Bureau Chief Executive Randall Rothenberg, says he was worried that Explorer's proposed privacy features would block not just the collection of consumer data, but also the delivery of some Web advertisements themselves. He says the features "seemed to equate the delivery of advertisements with privacy violations." He was especially troubled, he says, by the prospect of Microsoft turning the features on for all consumers, by default.

One other consideration: Some Microsoft executives were concerned that the preset-privacy plan might jeopardize support among ad-industry organizations that Microsoft wanted to rally against a proposed advertising deal between Google and Yahoo Inc., says a former Microsoft executive. A Microsoft spokeswoman declined to comment on that issue. U.S. regulators ended up blocking the deal.

The former Microsoft executive says he had never before experienced a debate at Microsoft "so driven by external influences and conflicting priorities to protect users" as the tussle over the Explorer privacy controls.

"It was a healthy debate," says Mr. Smith, the general counsel, with "well-informed views by people who are passionate."

When Microsoft released the browser in its final form in March 2009, the privacy features were a lot different from what its planners had envisioned. Internet Explorer required the consumer to turn on the feature that blocks tracking by websites, called InPrivate Filtering. It wasn't activated automatically.

What's more, even if consumers turn the feature on, Microsoft designed the browser so InPrivate Filtering doesn't stay on permanently. Users must activate the privacy setting every time they start up the browser.

Microsoft dropped another proposed feature, known as InPrivate Subscriptions, that would have let users further conceal their online browsing habits, by automatically blocking Web addresses suspected of consumer tracking if those addresses appeared on "black lists" compiled by privacy groups.

Mr. Cullen, Microsoft's chief privacy strategist, says the input of outsiders helped Microsoft strike a balance between privacy and advertising interests. The browser, he says, "was a better product than when it came off the drawing-room floor of the Internet Explorer group."

Advertising groups say they were pleased, too. "They ended up with something pretty excellent," says Mr. Rothenberg of the Interactive Advertising Bureau.

Published Aug. 2, 2010.

THE WALL STREET JOURNAL.
WSJ.com

On the Web's Cutting Edge, Anonymity in Name Only

BY EMILY STEEL and JULIA ANGWIN

You may not know a company called [x+1] Inc., but it may well know a lot about you.

From a single click on a website, [x+1] correctly identified Carrie Isaac as a young Colorado Springs parent who lives on about $50,000 a year, shops at Wal-Mart and rents kids' videos. The company deduced that Paul Boulifard, a Nashville architect, is childless, likes to travel and buys used cars. And [x+1] determined that Thomas Burney, a Colorado building contractor, is a skier with a college degree and looks like he has good credit.

The company didn't get every detail correct. But its ability to make snap assessments of individuals is accurate enough that Capital One Financial Corp. uses [x+1]'s calculations to instantly decide which credit cards to show first-time visitors to its website.

In short: Websites are gaining the ability to decide whether or not you'd be a good customer, before you tell them a single thing about yourself.

The technology reaches beyond the personalization familiar on sites like Amazon.com, which uses its own in-house data on its customers to show them new items they might like.

By contrast, firms like [x+1] tap into vast databases of people's online behavior—mainly gathered surreptitiously by tracking technologies that have become ubiquitous on websites across the Internet. They don't have people's names, but cross-reference that data with records of home ownership, family income, marital status and favorite restaurants, among other things. Then, using statistical analysis, they start to make assumptions about the proclivities of individual Web surfers.

"We never don't know anything about someone," says John Nardone, [x+1]'s chief executive.

Capital One says it doesn't use the full array of [x+1]'s targeting technology, and it doesn't prevent people from applying for any card they want. "While we suggest products that we believe will be of interest to our visitors, we do not limit their ability to easily explore all products available," spokeswoman Pam Girardo says.

A Wall Street Journal investigation into online privacy has found that the analytical skill of data handlers like [x+1] is transforming the Internet into a place where people are becoming anonymous in name only. The findings offer an early glimpse of a new, personalized Internet where sites have the ability to adjust many things—look, content, prices—based on the kind of person they think you are.

New York-based Demdex Inc., for instance, helps websites build "behavioral data banks" that tap sources including online-browsing records, retail purchases and a database predicting a person's spot in a corporate hierarchy. It crunches the data to help retailers customize their sites to target the person they think is visiting.

"If we've identified a visitor as a midlife-crisis male," says Demdex CEO Randy Nicolau, a client, such as an auto retailer, can "give him a different experience than a young mother with a new family." The guy sees a red convertible, the mom a minivan.

The technology raises the prospect that different visitors to a website could see different prices as well. Price discrimination is generally legal, so long as it's not based on race, gender or geography, which can be deemed "redlining."

In financial services, fair-lending laws prohibit discrimination based on race, color, religion, national origin, gender, receipt of public assistance or marital status. The laws also require that borrowers have access to any data used to evaluate their creditworthiness.

But the law doesn't specifically bar using web-browsing history to make lending decisions. That means, in theory, a bank could deny a loan based on knowledge of the applicant's visits to, say, gambling sites. In such a case, however, the bank would be required to let the applicant see the browsing data and correct it if inaccurate.

Capital One says it doesn't use [x+1] or browsing history in lending decisions. Rather, it uses [x+1] to suggest products to individuals.

The regulators who monitor fair lending at the Federal Trade Commission say suggesting offers isn't illegal. But it could violate the law if the suggestions result in protected groups such as minorities being steered into paying higher credit-card rates despite having solid credit.

"Steering can be a law violation depending on how they do it," says Alice Hrdy, an assistant director at the FTC. "Credit decisions have to be based on the customer's creditworthiness."

Capital One spokeswoman Ms. Girardo says, "Our practices are fully compliant with banking regulations and privacy laws."

[x+1] says none of its credit-card services use gender, ethnicity or age data. It adds that the company doesn't have the names of the individuals it analyzes.

The idea of using data about website visitors' offline lives was controversial in 1999 when [x+1] first started a website-personalization business. The FTC was investigating the privacy implications of online-advertising network DoubleClick Inc.'s acquisition of Abacus Direct, which tracked people's "offline" purchases at traditional retailers.

After a flood of negative publicity, DoubleClick agreed not to combine its online data with Abacus's offline data. For years, DoubleClick's experience deterred other companies from merging online and offline data.

[x+1] struggled to stay afloat. It cycled through six CEOs and three names, including Poindexter Systems (after a nerd scientist in "Felix the Cat" cartoons).

In 2008, Mr. Nardone took the helm just as things changed. Online ad spending rebounded and marketplaces for online data sprang up, letting companies like his tap data about people's Web browsing. Traditional data brokers (mostly serving direct-mail and catalog companies) began making data available online, too, but with names stripped out to address privacy worries.

Mr. Nardone saw opportunity. He revived the company's decade-old data-crunching patent for a "predictive optimization engine," now turbocharged with newly available data. "I discovered very quickly that we were going back to the original roots of the company," he says.

He found a receptive audience for site customization in the credit-card business, nabbing Capital One. [x+1] says clients pay $30,000 to $200,000 a month for its technology.

"You don't get information on everybody, but there are ways of doing analysis that you can fill out the gaps" says Ted Shergalis, [x+1]'s co-founder and chief strategy officer. "That is the whole science of this."

Its technology works like this: A visitor lands onCapital One's credit-card page, and [x+1] instantly scans the information passed between the person's computer and the web page, which can be thousands of lines of code containing details on the user's computer. [x+1] also uses a new service from Digital Envoy Inc. that can determine the ZIP code where that computer is physically located. For some clients (but not Capital One), [x+1] also taps additional databases of web-browsing history.

Armed with its data, [x+1] taps consumer researcher Nielsen Co. to assign the visitor to one of 66 demographic groups.

In a fifth of a second, [x+1] says it can access and analyze thousands of pieces of information about a single user. It quickly scans for similar types of Capital One customers to make an educated guess about which credit cards to show the visitor.

To gauge the system's accuracy, the Journal asked eight people to visit the credit-card page of Capital One's site and note the credit cards they were shown. The Journal also analyzed the computer code that zipped back and forth between the testers' computers and Capital One.

Separately, the Journal asked its testers to click on a custom website that [x+1] built to demonstrate its technology. After the testers clicked on that site, [x+1] described to the Journal what it knew about each person.

Throughout both of these processes, the testers didn't reveal any personal information.

[x+1]'s assessments of the testers were generally accurate, though some specific details missed the mark. For instance, [x+1] correctly placed Ms. Isaac, the Colorado Springs mom, in a Nielsen demographic segment called "White Picket Fences." People in this group live in small cities, have a median household income of $53,901, are 25 to 44 years old with kids, work in white-collar or service jobs, generally own their own home, and have some college education.

All of those points were correct for Ms. Isaac—to her surprise. "They pinpointed my income more accurately than I remembered it," she says.

But the "White Picket Fence" category wasn't 100% accurate. It suggested Ms. Isaac might read People en Espanol, watch Toon Disney and drive a Nissan Frontier truck. In fact, she doesn't speak Spanish, doesn't subscribe to cable TV and doesn't drive a truck.

Nielsen says its segments are intended to provide a broad framework to help marketers understand their customers, rather than an exact template.

[x+1] says its analysis isn't meant to be pinpoint accurate, either. "It is just saying, 'Do I have better than 50-50 odds of guessing what this anonymous user is going to want?'" says Mr. Shergalis, the company co-founder.

The Journal also captured and analyzed 5,219 lines of code that passed between Ms. Isaac's computer and Capital One. That code contained some of the results of [x+1]'s analysis, which was generally accurate, putting her in "Colorado Springs" with a "midscale" income and saying she was either a college grad or had "some college."

As a result of the analysis, Capital One showed her some of its least generous cards, which it describes as being for people with "average" credit. The bank defines "average" as people who might not currently have a card, or whose credit limits are below $5,000 or who "may have been late" on a loan or card in the past six months.

Ms. Isaac says that category fits. She and her husband use only debit cards after using credit cards in the past.

Capital One's Ms. Girardo says: "Like every marketer, online and offline, we're making an educated guess about what we think consumers will like and they are free to choose another product of their liking."

[x+1] zeroed in on Mr. Boulifard's love of travel. The Nashville architect was shown only one card on the Capital One website: a "VentureOne Rewards" card, shown floating above a beach scene, with a headline: "Still searching? Get double miles with Venture."

Mr. Boulifard is a member of several frequent-flier clubs and is saving up for a Europe trip. He routinely shops for hotels online and uses his American Express card to rack up travel points. However, Capital One's suggestion didn't tempt him to switch. "I have 90,000 points on my American Express, so I'm going to Europe on that," he says.

With Teresa Britton, [x+1]'s algorithms bumped into the limits of their ability to capture some of the complexities of modern America. The company correctly pegged her as a Greensboro, N.C., resident and assigned her to the "Young Influentials" Nielsen segment, a group that lives in suburbs and earns about $50,000 a year.

That underestimates her income, Ms. Britton says. But she really bristled at Nielsen's suggestions that she might buy rap music or read Vibe, the hip-hop magazine. Ms. Britton is white, and her husband is black.

"I don't know if they somehow got me and my husband mixed," she says. That said, her husband doesn't read Vibe or buy much rap music. "That is so stereotypical," she says. [x+1] says its Nielsen data are from broad segments and "don't apply at the individual level."

The technology was better at sizing up Mr. Burney, the Colorado contractor who builds ski-vacation homes. He saw only one credit card, the Capital One Prestige Platinum, under a headline: "Our best rewards at a glance." The pitch included an initial 0% interest rate and no annual fee.

That card made sense, Mr. Burney says. Not only does he tend to charge a lot of expenses for his construction business, but "I have a wallet full of platinum credit cards," he says. "My credit is sparkling."

Based on Mr. Burney's visit to [x+1]'s test site, the company pegged him as a member of a Nielsen segment called "God's Country." People in that group live in small towns or rural areas, have median household income of $86,724, are 35 to 54 years old with no kids, work in management, mostly own their homes and are college graduates.

Mr. Burney is, in fact, a homeowner, a college grad and a manager, and has no kids. At 28, he's younger than predicted, and his income is less than predicted.

When he saw the 3,748 lines of code that passed in an instant between his computer and Capital One's website, Mr. Burney said: "There's a shocking amount of information there." Buried in the code were references to his income level ("uppermid"), education ("college%252b graduate") and his town ("avon").

In fact, [x+1]'s assessment of Mr. Burney's location and Nielsen demographic segment are specific enough that it comes extremely close to identifying him as an individual—that is, "de-anonymizing" him—according to Peter Eckersley, staff scientist at the Electronic Frontier Foundation, a privacy-advocacy group.

Mr. Eckersley does research in the field of de-anonymization, the mathematics of identifying individuals based on a few specific details from their life. In the jargon of the field, Mr. Eckersley says, all that's needed to uniquely identify one person is a total of 33 "bits" of information about him or her.

Calculating "bits" gets complex, as some facts about a person are more valuable—and thus have more "bits"—than others. ZIP codes and birthdates, for instance, are extremely valuable when zeroing in on individuals.

Bottom line: Mr. Eckersley determined Mr. Burney's location (the small town of Avon, Colo.) and his Nielsen demographic segment ("God's Country") together offered about 26.5 bits of information that could be used to identify Mr. Burney individually.

That's enough to narrow him down to one of just 64 or so people worldwide.

With one more piece of information about him, such as his age, Mr. Eckersley says, it's likely that Mr. Burney could be de- anonymized. "You're starting to look very close to identified."

Mr. Nardone of [x+1] acknowledges the possibility of de-anonymization, but says it isn't worth the effort: The company already has enough information to sell. "It would be a massive undertaking," he says, "and it is hard enough to make money."

Published Aug. 4, 2010. The article has been updated to incorporate a correction published Aug. 13, 2010.

For interactive graphics related to this story, click this link.

THE WALL STREET JOURNAL.
WSJ.com

Stalking by Cellphone

BY JUSTIN SCHECK

Phone companies know where their customers' cellphones are, often within a radius of less than 100 feet. That tracking technology has rescued lost drivers, helped authorities find kidnap victims and let parents keep tabs on their kids.

But the technology isn't always used the way the phone company intends.

One morning in the summer of 2009, Glenn Helwig threw his then-wife to the floor of their bedroom in Corpus Christi, Texas, she alleged in police reports. She packed her 1995 Hyundai and drove to a friend's home, she recalled recently. She didn't expect him to find her.

The day after she arrived, she says, her husband "all of a sudden showed up." According to police reports, he barged in and knocked her to the floor, then took off with her car.

The police say in a report that Mr. Helwig found his wife using a service offered by his cellular carrier, which enabled him to follow her movements through the global-positioning-system chip contained in her cellphone.

Mr. Helwig, in an interview, acknowledged using the service to track his wife on some occasions. He says he signed up for the tracking service in 2009. "AT&T had this little deal where you could find your family member through her cellphone," he says. But he didn't use it to find his wife that day, he says. Mr. Helwig, who is awaiting trial on related assault charges, declined to comment further about the matter. He has pleaded not guilty.

The allegations are a stark reminder of a largely hidden cost from the proliferation of sophisticated tracking technology in everyday life—a loss of privacy.

Global-positioning systems, called GPS, and other technologies used by phone companies have unexpectedly made it easier for abusers to track their victims. A U.S. Justice Department report in 2009 estimated that more than 25,000 adults in the U.S. are victims of GPS stalking annually, including by cellphone.

In the online world, consumers who surf the Internet unintentionally surrender all kinds of personal information to marketing firms that use invisible tracking technology to monitor online activity. A Wall Street Journal investigation of the 50 most-popular U.S. websites found that most are placing intrusive

tracking technologies on the computers of visitors—in some cases, more than 100 tracking tools at a time.

The cellphone industry says location-tracking programs are meant to provide a useful service to families, and that most providers take steps to prevent abuse. Mike Altschul, chief counsel for wireless-telecommunications trade group CTIA, says recommended "best practices" for providers of such services include providing notification to the person being tracked.

Mr. Helwig's wife had received such a notification, by text message, from AT&T. A spokesman for AT&T Inc. says it notifies all phone users when tracking functions are activated. But users don't have the right to refuse to be tracked by the account holder. Turning off the phone stops the tracking.

Cellphone companies will deactivate a tracking function if law-enforcement officials inform them it is being used for stalking. Mr. Altschul says authorities haven't asked carriers to change their programs. He adds that carriers have long supported programs to give untraceable cellphones to domestic-violence victims.

In Arizona in 2010, Andre Leteve used the GPS in his wife's cellphone to stalk her, according to his wife's lawyer, Robert Jensen, before allegedly murdering their two children and shooting himself. Mr. Jensen says Mr. Leteve's wife, Laurie Leteve, didn't know she was being tracked until she looked at one of the family's monthly cellphone bills, more than 30 days after the tracking began. Mr. Leteve, a real-estate agent, is expected to recover. He has pleaded not guilty to murder charges, and is awaiting trial. The law firm representing him declined to comment.

In a suspected murder-suicide in 2009 near Seattle, a mechanic named James Harrison allegedly tracked his wife's cellphone to a store. After he found her there with another man, he shot to death his five children and himself, according to the Pierce County Sheriff's Office.

Therapists who work with domestic-abuse victims say they are increasingly seeing clients who have been stalked via their phones. At the Next Door Solutions for Battered Women shelter in San Jose, Calif., director Kathleen Krenek says women frequently arrive with the same complaint: "He knows where I am all the time, and I can't figure out how he's tracking me."

In such cases, Ms. Krenek says, the abuser is usually tracking a victim's cellphone. That comes as a shock to many stalking victims, she says, who often believe that carrying a phone makes them safer because they can call 911 if they're attacked.

There are various technologies for tracking a person's phone, and with the fast growth in smartphones, new ones come along frequently. In 2010, researchers with iSec Partners, a cyber-security firm, described in a report how anyone could track a phone within a tight radius. All that is required is the target person's cellphone number, a computer and some knowledge of how cellular networks work, said the report, which aimed to spotlight a security vulnerability.

The result, says iSec researcher Don Bailey, is that "guys like me, who shouldn't have access to your location, have it for very, very, very cheap."

That is, in part, an unintended consequence of federal regulations that require cellphone makers to install GPS chips or other location technology in nearly all phones. The Federal Communications Commission required U.S. cellular providers to make at least 95% of the phones in their networks traceable by satellite or other technologies by the end of 2005. The agency's intention was to make it easier for people in emergencies to get help. GPS chips send signals to satellites that enable police and rescue workers to locate a person.

To a large extent, that potential has been fulfilled. In 2009, for example, police in Athol, Mass., working with a cellphone carrier, were able to pinpoint the location of a 9-year-old girl who allegedly had been kidnapped and taken to Virginia by her grandmother. In December 2009, police in Wickliffe, Ohio, tracked down and arrested a man who allegedly had robbed a Pizza Hut at gunpoint by tracking the location of a cellphone they say he had stolen.

Mr. Altschul, of the cellphone-industry trade group, says the tracking technology has been of great help to both law-enforcement officials and parents. "The technology here is neutral," he says. "It's actually used for peace of mind."

But as GPS phones proliferated, tech companies found other uses for the tracking data. Software called MobileSpy can "silently record text messages, GPS locations and call details" on iPhones, BlackBerrys and Android phones, according to the program's maker, Retina-X Studios LLC. For $99.97 a year, a person can load MobileSpy onto someone's cellphone and track that phone's location.

Craig Thompson, Retina-X's operations director, says the software is meant to allow parents to track their kids and companies to keep tabs on phones their employees use. He says the company has sold 60,000 copies of MobileSpy. The company sometimes gets calls from people who complain they are being improperly tracked, he says, but it hasn't been able to verify any of the complaints.

Installing such programs requires a person to physically get hold of the phone to download software onto it.

GPS-tracking systems provided by cellular carriers such as AT&T and Verizon Communications Inc. are activated remotely, by the carriers.

Domestic-violence shelters have learned the consequences. As soon as victims arrive at shelters run by A Safe Place, "we literally take their phones apart and put them in a plastic bag" to disable the tracking systems, says Marsie Silvestro, director of the Portsmouth, N.H., organization, which houses domestic-violence victims in secret locations so their abusers can't find them.

The organization put that policy in place after a close call. On Feb. 26, 2010, Jennie Barnes arrived at a shelter to escape her husband, Michael Barnes, according to a police affidavit filed in a domestic-violence case against Mr.

Barnes in New Hampshire state court. Ms. Barnes told police she was afraid that Mr. Barnes, who has admitted in court to assaulting his wife, would assault her again.

Ms. Barnes told a police officer that "she was in fear for her life," according to court filings. The next day, a judge issued a restraining order requiring Mr. Barnes to stay away from his wife.

Later that day, court records indicate, Mr. Barnes called his wife's cellular carrier, AT&T, and activated a service that let him track his wife's location. Mr. Barnes, court records say, told his brother that he planned to find Ms. Barnes.

The cellular carrier sent Ms. Barnes a text message telling her the tracking service had been activated, and police intercepted her husband. Mr. Barnes, who pleaded guilty to assaulting his wife and to violating a restraining order by tracking her with the cellphone, was sentenced to 12 months in jail. A lawyer for Mr. Barnes didn't return calls seeking comment.

Another source for cellphone tracking information: systems meant to help police and firefighters. Some cellular carriers provide services for law-enforcement officers to track people in emergencies. Using such systems requires a person to visit a special website or dial a hot-line number set up by the carrier and claim the data request is for law-enforcement purposes.

Cellular carriers say they try to verify that callers are legitimate. An AT&T spokesman says an office is manned around the clock by operators who ask for subpoenas from law-enforcement officials using the system.

But federal law allows carriers to turn over data in emergencies without subpoenas. Al Gidari, a lawyer who represents carriers such as Verizon, says such location-tracking systems can be easy to abuse. Police, he says, often claim they need data immediately for an emergency like a kidnapping, and therefore don't have time to obtain a warrant, in which a judge must approve an information request.

In Minnesota, Sarah Jean Mann claimed in 2009 in a county-court petition for a restraining order that her estranged boyfriend, a state narcotics agent, followed her by tracking her cellphone and accessing her call and location records through such a system. The court issued the restraining order. The boyfriend, Randy Olson, has since resigned from the police force. He didn't respond to calls seeking comment.

Mr. Gidari says law-enforcement's easy access to such data makes the systems easy to abuse. He says carriers would like to have a system in place requiring agents to get warrants. Without such a requirement, there is little carriers can do to resist warrantless requests, say Mr. Gidari and Mr. Altschul of trade group CTIA. Federal law says carriers may comply with such requests, and law-enforcement agencies have pressured them to maintain the tracking systems, Mr. Gidari says.

The easiest way for stalkers to locate a target—and perhaps the most common, say therapists who work with victims and abusers—is by using systems offered by carriers. When cellphone users sign up for a "family plan" that includes two or more phones, they have the option to contact the carrier and activate a tracking feature intended to allow them to keep tabs on their children.

The AT&T FamilyMap program, for example, is free for 30 days and requires only a phone call to activate. "Know where your kids and loved ones are at any time!" says AT&T's website. The system is for parents, says an AT&T spokesman. He says the company hasn't received complaints about FamilyMap being used by stalkers.

The system provides an on-screen map on the smartphone or computer of the person doing the tracking. A dot on the map shows the location and movement of the person being followed. The carrier sends a text-message to the person being tracked that the phone is registered in the program.

These add-on services can be lucrative for carriers. AT&T debuted its FamilyMap system in April 2009. It charges $9.99 a month to track up to two phones, $14.99 for up to five. FamilyMap users must agree to "terms-of-use" stating that they may not use the system to "harass, stalk, threaten" or otherwise harm anyone.

In Corpus Christi, Mr. Helwig and his wife, who had been married since early 2008, bought phones under an AT&T family plan. Mr. Helwig says he activated the feature in 2009. His wife says she received a text message that a tracking function had been activated on her phone, but wasn't sure how it was activated. Her husband, she says, initially denied turning on the tracking function.

She says she eventually came up with a plan to flee to the house of a family whose children she baby-sat. Her husband "had no idea where they lived" or even their names, she says. As she was packing, her husband confronted her. They argued, and, according to her statements in police reports, Mr. Helwig dragged her around by her hair.

The police came. She says she told them she didn't want them to arrest Mr. Helwig, that she simply wanted to leave. The police told Mr. Helwig to stay away from her for 24 hours, she says.

As she drove to her friend's house, she says, she made sure her phone was off so Mr. Helwig couldn't track her. But she turned it on several times to make calls. The next day, Mr. Helwig was outside in a rage, according to police reports.

Mr. Helwig forced his way into the house, pushed her to the floor, took her car keys and drove away in her Hyundai, according to police reports.

Police arrested Mr. Helwig a short distance away. Mr. Helwig, a firefighter, is facing charges of assault and interfering with an emergency call. His trial is scheduled to begin this summer.

Mr. Helwig and his wife divorced, and she left Corpus Christi. She says she doesn't want to testify against him. She says she is more careful about trusting her cellphone now.

Published Aug. 5, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Google Agonizes on Privacy as Ad World Vaults Ahead

BY JESSICA E. VASCELLARO

A confidential, seven-page Google Inc. "vision statement" shows the information-age giant in a deep round of soul-searching over a basic question: How far should it go in profiting from its crown jewels—the vast trove of data it possesses about people's activities?

Should it tap more of what it knows about Gmail users? Should it build a vast "trading platform" for buying and selling Web data? Should it let people pay to not see any ads at all?

These and other ideas big and small—the third one was listed under "wacky"—are discussed in the document, which was reviewed by The Wall Street Journal and compiled in late 2008 by Aitan Weinberg, now a senior product manager for interest-based advertising. Along with interviews with more than a dozen current and former employees, the vision statement offers a candid, introspective look at Google's fight to remain at the vanguard of the information economy.

Google is pushing into uncharted privacy territory for the company. Until recently, it refrained from aggressively cashing in on its own data about Internet users, fearing a backlash. But the rapid emergence of scrappy rivals who track people's online activities and sell that data, along with Facebook Inc.'s growth, is forcing a shift.

A person familiar with the matter called the vision statement a "brainstorming document" and said it wasn't presented to senior executives. Some of its ideas are "complete non-starters," this person said. Efforts to reach Mr. Weinberg weren't successful.

Still, several have been implemented. Among them: In 2009, Google for the first time started collecting a new type of data about the websites people visit, using it to track and show them ads across the Internet.

Worries about the size of Google's data cache are "hypothetical," said co-founder Larry Page in a 2010 response to a reporter's question about privacy. "It is always easy to be fearful of what could happen, right?"

As Google changes, it is likely to bring the rest of the online world with it. With more users than any other Internet company, it has an unparalleled ability to make new ad-targeting methods mainstream. The company also actively participates in trade groups that regularly craft new privacy practices among themselves in hopes of thwarting legislation. The Federal Trade Commission said in 2009 that the field can regulate itself as long as companies disclose their practices to users, among other things.

Google is overwhelmingly important to online privacy. Roughly 75% of global Internet users, or 943.8 million people, used its services in June 2010, more than any other Web company, according to comScore.

The vision statement describes the company's immense search database as "the BEST source of user interests found on the Internet," during a discussion of ways to make ads more relevant to users. "No other player could compete," it says. Later, the document warns that some ideas range from "safe" to "not" safe.

The most aggressive ideas would put Google at the cutting edge of the business of tracking people online to profit from their actions. A data-trading marketplace, for instance, would allow personal information from many sources—including Google—to be combined and used for highly personalized tracking of individuals.

Tiny companies like BlueKai Inc. and eXelate Media Ltd. already offer some of these services, pressuring Google to match them. A Wall Street Journal investigation, "What They Know," is examining the widening trade in this kind of data and the consequences for individual privacy.

Google trails in some of these techniques by choice. Famous for its unofficial corporate motto, "Don't Be Evil," for years it resisted using any method to track people online without their knowledge at the fierce insistence of founders Sergey Brin and Mr. Page. But the two men have gradually decided they can begin exploiting the data their company controls, without exploiting consumers, according to interviews with more than a dozen current and former employees.

The founders believe they are improving the Internet user's experience, said Alma Whitten, who leads Google's privacy engineering, in a June 2010 interview. "What's good for the consumer is good for the advertiser."

A recent Journal examination of the proliferation of online tracking found that Google's tracking code appeared on 45 of the 50 most popular U.S. websites. (For more details on those findings, go to WSJ.com/WTK.)

The 2008 vision statement along with a dozen other internal documents reviewed by the Journal tell the inside story of how Google dragged its feet while its founders' views evolved.

Selling ads is Google's big money-maker, but the online-ad business is broadening away from Google's sweet spot, selling ads tied to the search-engine terms people use. Instead, advertisers want to target people based on more

specific personal information such as hobbies, income, illnesses or circles of friends.

The changes at Google reflect a power realignment online. For years, the strongest companies on the Internet were the ones with the most visitor traffic. Today, the power resides with those that have the richest data and are the savviest about using it.

That has propelled Internet ad companies into an arms race so swift that even Google fears being left behind. One slide from an internal presentation in mid-2008, which was reviewed by the Journal, is headlined bluntly: "Get in the Game."

That particular slide describes the importance of breaking into the lucrative business of selling "display" ads, which are larger ads with pictures, as opposed to smaller text ads. Today, Google still trails market leader Yahoo in U.S. display-ad revenue, according to analysts.

Google still leads the Internet pack overall, of course. Its revenue, $23.7 billion in 2009, is more than three times Yahoo's, its closest competitor. Its online advertising business is growing faster than those of its publicly held U.S. rivals.

But Google's revenue growth has slowed dramatically. And social-networking powerhouse Facebook is a widening threat with its ability to sell highly targeted ads to its more than 500 million users.

Facebook fears run deep at Google, which is designing its own social-networking service. In a sign of how quickly things change, the 2008 vision statement scarcely mentioned social networks.

Google also plans to go head-to-head with Facebook's "Like" button—a tiny tool on many websites that lets people tell friends they "like" something. Each click gives Facebook valuable, personal data about people's interests.

Few online companies have the potential to know as much about its users as Google. Consider 26-year-old Ari Brand, an actor living in Manhattan's East Village. Google has access to the fact he paid $733 for a flat-screen TV, because he uploaded his budget to Google Docs, an online word processor and spreadsheet. It has access to the 23,000 emails he has sent through Gmail. Google also saves searches tied to the network address of Mr. Brand's computer, which it makes anonymous after 18 months.

Significantly, however, Google doesn't mix those separate pots of personal data. For instance, it doesn't use data gleaned from a person's Gmail account to target ads to that person elsewhere online. Google's computers do, however, scan Gmail messages to place contextual ads next to the emails themselves.

Google also says much of its data can't be tied to a person by name.

Executives long considered the privacy risks too great relative to the business rewards. According to people familiar with Google's thinking, they felt the company was being held to a higher standard than less well-known firms and preferred to let more aggressive rivals test the boundaries.

Concerns about antitrust scrutiny also heightened the risk of finding new ways to profit from Google's exclusive data.

As recently as 2006 or so, Google's sights weren't set on Facebook—they were set on AOL and Yahoo, which together controlled roughly 40% of the U.S. display-ad business, analysts say.

One big obstacle in winning more of that business was Google co-founder Mr. Page, who objected to letting Google's advertising customers work with companies that installed "cookies" on people's computers for purposes of serving ads and tracking their performance. Cookies are little text files that can, among other things, be used to help track people's activities online to show them ads targeted to their interests.

Those policies hurt Google's display ad sales because the company wouldn't let advertisers use technology they were used to. Google didn't use ad-targeting cookies itself, either. That meant Google could sell ads based only on the name or content of a page—for instance, putting a shoe ad on a page about shoes. That is known as "contextual" targeting, and many advertisers consider it less effective than "behavioral" targeting, which identifies specific users and their interests.

In 2006, Gokul Rajaram, then a senior Google staffer, and ad-sales executive Tim Armstrong tried to change Mr. Page's mind about letting other companies place cookies.

In an interview, Mr. Rajaram recalls that he thought it would be an easy sell. A growing number of advertisers were refusing to buy display ads from Google. Market research showed AOL and Yahoo were trouncing Google in the display market.

Messrs. Page and Brin weren't swayed. "I was kind of shocked," Mr. Rajaram says. "They just didn't look at it the same way."

As factions inside Google fought over the issue, an opportunity arose. DoubleClick Inc., a giant in the business of placing display ads on websites, put itself up for sale—and Google archrival Microsoft Corp. was circling.

Google executives were leery of the way DoubleClick used cookies to track people online, on the principle that many users had no idea they were being tracked, people familiar with the situation say.

But an acquisition of DoubleClick would instantly bring in display-advertising expertise and clients, they thought.

In 2007, Google agreed to buy DoubleClick for $3.1 billion. At the time, some employees joked Google had to spend billions just to get Mr. Page to like cookies, people familiar with the matter say.

Google and DoubleClick executives huddled to decide how to blend the two companies' products. They had a lot of ground to make up.

According to a resulting presentation slide, dated July 2008—the one headlined "Get in the Game"—Google offered fewer ways to measure an ad's

effectiveness than Atlas, a rival owned by Microsoft. And Google had none of the behavioral-targeting capacities of AOL's Tacoda unit—meaning it couldn't target ads to people based on websites they visited.

Google executives finally agreed it was cookie time. As a result, every page where Google sold a display ad began installing a DoubleClick cookie on users' computers.

For the first time, Google had the ability to deliver ads targeted to individual people's computers. But just because it had the ability, Google didn't start using it. There was still too much internal resistance.

Mr. Weinberg, the author of the 2008 "vision statement," came to Google from DoubleClick. He and a small group of product managers and marketing officials began discussing the ways Google could target ads to people more aggressively.

His memo, stamped "INTERNAL CONFIDENTIAL," acknowledged the delicateness of the subject. Audience targeting is "of a sensitive nature," it stated in the very first sentence, due to the possibility of "mis-understanding" among users.

The memo then went on to outline a sweeping vision in which Google could get other websites from around the Internet to share their data with it for the purpose of targeting ads.

The document also says Google could start selling ads across the Web based on the things it knew about people from their Gmail accounts, and also from their use of Google's Checkout service, a PayPal rival.

All of that would be a significant change. Currently, although Google places contextual ads within a user's Gmail account, it doesn't follow that person to other websites with those ads.

The document shows awareness of the privacy implications. Nothing would happen "without strong consideration of privacy, legal and industry best practices in mind," it states. A goal should be to limit users' feeling of "creepiness" from seeing finely targeted ads, it says.

By late 2008, Google executives were preparing to launch ads targeted at users' interests. But the specifics still remained controversial.

Tensions erupted during a meeting with about a dozen executives at Google's Mountain View, Calif., headquarters about 18 months ago when Messrs. Page and Brin shouted at each other over how aggressively Google should move into targeting, according to a person who had knowledge of the meeting. "It was awkward," this person said. "It was like watching your parents fight."

Mr. Brin was more reluctant than Mr. Page, this person said. Eventually, he acquiesced and plans for Google to sell ads targeted to people's interests went ahead.

Google launched the new advertising product, "interest-based ads" in March 2009. The service, currently available only to a limited group of advertisers, uses cookies to track any time a user visits one of the more than one million sites where Google sells display ads.

To offset the founders' concerns about cookies' secretiveness, Google set up a page, www.google.com/ads/preferences, where people can opt out and see what Google has inferred about their interests.

Google adopted other vision-statement ideas. In September 2010, it launched its new ad exchange, which lets advertisers target individual people—consumers in the market for shoes, for instance—and buy access to them in real time as they surf the Web. Google takes a cut of each ad sale.

In short, Google is trying to establish itself as the clearinghouse for as many ad transactions as possible, even when those deals don't actually involve consumer data that Google provides or sees.

The further step in that progression would be for Google to become a clearinghouse for everyone's data, too. That idea, also laid out in the vision statement, is still being considered, people familiar with the talks say. That would put Google—already one of the biggest repositories of consumer data anywhere—at the center of the trade in other people's data as well.

For excerpts from Google's Confidential “Vision Statement,” click this link: http://online.wsj.com/article/SB10001424052748704164904575421103515121436.html

Published Aug. 10, 2010.

THE WALL STREET JOURNAL.
WSJ.com

On the Web, Children Face Intensive Tracking

BY STEVE STECKLOW

A Wall Street Journal investigation into online privacy has found that popular children's websites install more tracking technologies on personal computers than do the top websites aimed at adults.

The Journal examined 50 sites popular with U.S. teens and children to see what tracking tools they installed on a test computer. As a group, the sites placed 4,123 "cookies," "beacons" and other pieces of tracking technology. That is 30% more than were found in an analysis of the 50 most popular U.S. sites overall, which are generally aimed at adults.

The most prolific site: Snazzyspace.com, which helps teens customize their social-networking pages, installed 248 tracking tools. Its operator described the site as a "hobby" and said the tracking tools come from advertisers.

Starfall.com, an education site for young children, installed the fewest, five.

The research is part of a Journal investigation into the expanding business of tracking people's activities online and selling details about their behavior and personal interests.

The tiny tracking tools are used by data-collection companies to follow people as they surf the Internet and to build profiles detailing their online activities, which advertisers and others buy. The profiles don't include names, but can include age, tastes, hobbies, shopping habits, race, likelihood to post comments and general location, such as city.

Selling the data is legal, but controversial, especially when it involves young people. Two companies identified by the Journal as selling teen data initially denied doing so. Only when shown evidence that they were offering data for sale—in one case, it was labeled "teeny boppers"—did they confirm it.

The Journal found that many popular children's sites are run by small companies or mom-and-pops, and privacy practices vary widely. Among the sites studied, the Journal identified one, y8.com—featuring kids' games with names like "Crush the Castle 2" and "Dreamy Nails Makeover"—that has had ties to a pornography site, xnxx.com, according to Internet registration records. Y8 installed 69 tracking files on the Journal's test computer. It also asks users to provide an email address to register.

"Children are safe on y8," a site employee named Olivier G. said in response to emailed questions. "We are *strongly against* the exposure of children to any adult content." Asked twice about y8.com's apparent ties to a pornography site, he didn't respond.

The Journal's study focused on sites popular with young people according to comScore Media Metrix.

Companies placing the tracking tools say the information they collect is anonymous and mainly used to deliver targeted ads or to gauge ads' effectiveness. They also say they don't collect "personally identifiable information" like names or email addresses and generally don't specifically target children.

Collecting data on minors is regulated, albeit lightly. The only federal restrictions require parental consent to collect names and other personal information of children under 13 in most circumstances. Currently, the Federal Trade Commission is considering whether changes to the law are warranted. No changes are expected before next year.

Many kids' sites are heavily dependent on advertising, which likely explains the presence of so many tracking tools. Research has shown children influence hundreds of billions of dollars in annual family purchases.

Google Inc. placed the most tracking files overall on the 50 sites examined. A Google spokesman said "a small proportion" of the files may be used to determine computer users' interests. He also said Google doesn't include "topics solely of interest to children" in its profiles.

Still, Google's "Ads Preferences" page (google.com/ads/preferences) displays what Google has determined about web users' interests. There, Google accurately identified a dozen pastimes of 10-year-old Jenna Maas—including pets, photography, "virtual worlds" and "online goodies" such as little animated graphics to decorate a website.

"It is a real eye opener," said Jenna's mother, Kate Maas, a schoolteacher in Charleston, S.C., viewing that data.

Jenna, now in fifth grade, said: "I don't like everyone knowing what I'm doing and stuff."

A Google spokesman said its preference lists are "based on anonymous browser activity. We don't know if it's one user or four using a particular browser, or who those users are." He said users can adjust the privacy settings on their browser or use the Ads Preferences page to limit data collection.

As part of the project, the Journal calculated an "exposure index" for each site, taking into account the number of trackers on the site and data-handling practices of those trackers. Snazzyspace.com ranked highest in exposing users to potentially aggressive tracking. A site owned by Viacom Inc., neopets.com, where kids can create make-believe "pets," had the highest exposure index of sites popular with children under 12.

Viacom's Nickelodeon TV network accounted for eight of the 50 sites in the survey. On average, the eight installed 81 tracking tools, close to the 82 average for all 50 sites. One, a games site called Shockwave.com, installed 146; another game-and-video site, nick.com, installed 92.

The vast majority of tracking files on Nickelodeon sites were installed by other firms, such as ad networks. A Nickelodeon official said those services "are collecting data on what users like to see and do based on their web behaviors and activities."

Many tools raise no privacy concerns. They might merely remember, say, where users pause in a game, so they aren't forced to start every time they visit.

But other tools are used to develop profiles of web-surfing behavior. Those can be used to deliver targeted ads that home in on children's concerns—say, dieting ads aimed at youngsters worried about their weight.

The number of tracking files installed by any specific site can vary from visit to visit. In the Journal's examination, the math-games site coolmath4kids.com installed 60 on a test computer.

However, when Angela La Fon, a teacher in Big Island, Va., checked her own computer with a tracker-detection tool called Abine, she found the site had installed 89 "cookies" on her machine. (Cookies are little text files that can give a computer a unique identity, which data-collection companies can use to track people's activities).

"That's creepy," says Ms. La Fon, who encouraged her six-year-old son, Lee, to use the site. "I wouldn't have thought I would have had 89 cookies, period. Much less than from one site."

Karen Davis, chief executive of coolmath4kids.com, declined to be interviewed, citing concern for her own privacy. In an email she wrote, "We are assured by our service providers that all data gathered is anonymous and compliant with all laws and privacy policies."

Several sites, including coolmath4kids.com, modified their privacy policies after being contacted by the Journal with its findings. For example, the math-games site no longer states that using cookies to collect anonymous data is "no big deal." Ms. Davis said she made the changes "to provide as much transparency as possible for our users."

A spokeswoman for weeworld.com, where kids can create a WeeMee avatar and chat with friends, said that as a result of a Journal analysis, it changed its privacy policy to provide a clearer explanation of how to disable cookies. Weeworld.com installed 144 tracking tools in the Journal's test.

Of the 50 sites examined by the Journal, only one had no posted privacy policy, the gaming site y8.com. Records at archive.org, a library of previous versions of websites, indicate that y8.com launched in the late 1990s as a sex site for adults at least 21 years old.

Y8.com became a game site aimed at a younger audience in 2006. ComScore reports that 12.2% of its users are 2 to 11 years old, and 22.8% are 12 to 17.

Internet registration records from December 2006 show that y8.com and a hard-core sex site, xnxx.com, shared the same mailing address in France, plus the same email address.

Later, the sites changed their contact information and no longer share the same addresses. On the website games.xnxx.com, which bills itself as offering "fun sex games," there is a prominent link at the top and bottom of the page to "non-adult" games on y8.com.

The y8.com employee, Olivier G., didn't respond to questions about who owns the site or its apparent relationship with xnxx.com. He wrote in an email that y8.com is "strongly against the collection and use of personal information." He also said "we don't do anything" with email addresses provided by users.

Parents hoping to let their kids use the Internet, while protecting them from snooping, are in a bind. That's because many sites put the onus on visitors to figure out how data companies use the information they collect.

Gaiaonline.com—where teens hang out together in a virtual world—says in its privacy policy that it "cannot control the activities" of other companies that install tracking files on its users' computers. It suggests that users consult the privacy policies of 11 different companies.

In a statement, gaiaonline.com said, "It is standard industry practice that advertisers and ad networks are bound by their own privacy policy, which is why we recommend that our users review those." The Journal's examination found that gaiaonline.com installed 131 tracking files from third parties, such as ad networks.

An executive at a company that installed several of those 131 files, eXelate Media Ltd., said in an email that his firm wasn't collecting or selling teen-related data. "We currently are not specifically capturing or promoting any 'teen' oriented segments for marketing purposes," wrote Mark S. Zagorski, eXelate's chief revenue officer.

But the Journal found that eXelate was offering data for sale on 5.9 million people it described as "Age: 13-17." In a later interview, Mr. Zagorski confirmed eXelate was selling teen data. He said it was a small part of its business and didn't include personal details such as names.

BlueKai Inc., which auctions data on Internet users, also said it wasn't offering for sale data on minors. "We are not selling data on kids," chief executive Omar Tawakol wrote in an email. "Let there be no doubt on what we do."

However, another data-collecting company, Lotame Solutions Inc., told the Journal that it was selling what it labeled "teeny bopper" data on kids age 13 to 19 via BlueKai's auctions. "If you log into BlueKai, you'll see 'teeny boppers' available for sale," said Eric L. Porres, Lotame's chief marketing officer.

Mr. Tawakol of BlueKai later confirmed the "teeny bopper" data had been for sale on BlueKai's exchange but no one had ever bought it. He said as a result of the Journal's inquiries, BlueKai had removed it.

The FTC is reviewing the only federal law that limits data collection about kids, the Children's Online Privacy Protection Act, or Coppa. That law requires sites aimed at children under 13 to obtain parental permission before collecting, using or disclosing a child's "personal information" such as name, home or email address, and phone and Social Security number. The law also applies to general-audience sites that knowingly collect personal information from kids.

The FTC is considering, among other things, whether to broaden "personal information" to include data "collected in connection with online behavioral advertising."

To try to avoid having to comply with Coppa, some sites state they prohibit kids under 13 from visiting. But that's easy for children to circumvent. Jenna Maas, the Charleston 10-year-old, opened an account on weeworld.com (which prohibits kids under 13 from registering) simply by fibbing about her age.

In Jenna's case, she got her mother's permission first. Ms. Maas says she lets Jenna visit sites for older kids "as long as I can monitor it."

Claire Quinn, weeworld.com's chief of safety, says the site has "tools in place" to prevent underage kids from joining, but "there is obviously no great age verification system out there."

FTC officials said website operators can't be held responsible if children lie about their age unless they glean from other information that a child is under 13.

Some experts believe the federal law also should apply to collecting data on teens, though not necessarily by requiring parental consent.

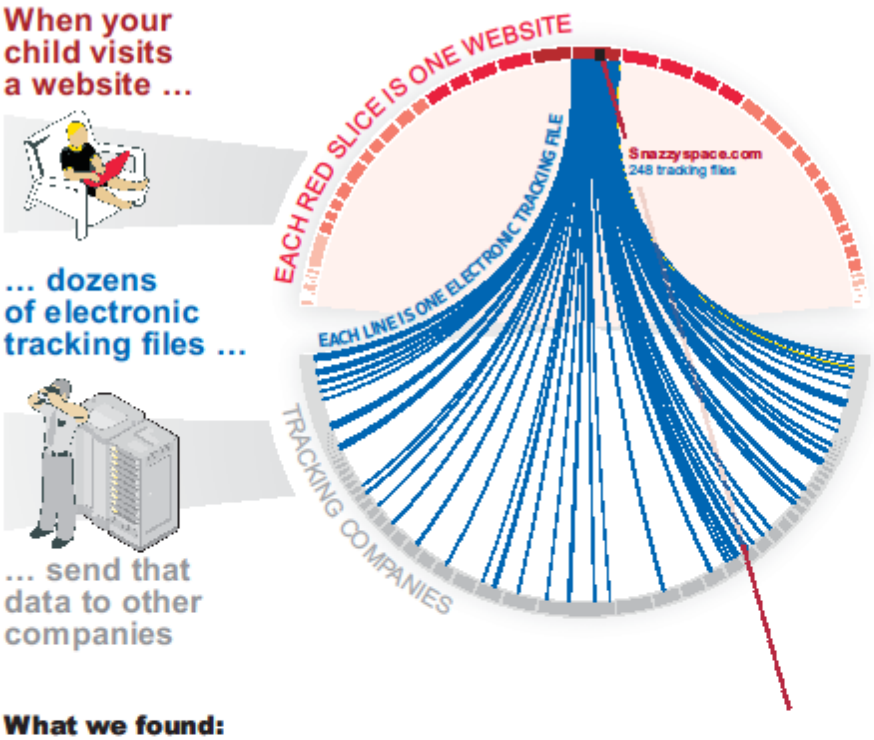
"We need clearer explanations of what's happening to their data online, that they can understand—not the kind of legalese in a privacy policy that basically obscures what's really going on," says Kathryn C. Montgomery, a professor of communication at American University.

Tom McGinty contributed to this report.

Published Sept. 18, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Explore the Data



What we found:
Among 50 of the Web's most-popular U.S. kids' sites, **Snazzyspace.com**, a site used by teens to decorate social-networking pages, hosted the most trackers and was least protective of its users' information.

How much each website exposes user data ...
The 'exposure index' was calculated based on a scoring of those tracking files and the total number of files on that site.

MOST EXPOSURE ⟷ LEAST EXPOSURE

SEE EXPOSURE FOR POPULAR KIDS AND TEEN SITES:

http://blogs.wsj.com/wtk-kids/

THE WALL STREET JOURNAL.
WSJ.com

How to Protect Your Child's Privacy Online

BY JENNIFER VALENTINO-DEVRIES

Websites popular among children and teens place more tracking technologies on users' computers than do the top websites aimed at adults, a Wall Street Journal investigation has found. But parents can take steps to limit their children's exposure.

Web-browsing activity is tracked by "cookies," "beacons" and "Flash cookies," small computer files or software programs installed on a computer when a user visits some Web pages. Some are useful. But others are used by companies to track users from site to site and build profiles of their online activities.

All Internet users, whether adults or children, can limit tracking by adjusting settings on Web browsers and Adobe Systems Inc.'s popular Flash program. These settings can delete cookies and limit what types of cookies may be placed on the computer. For additional protection, parents also can install small programs, called "add-ons," to a child's browser. And parents can prevent children from seeing behaviorally targeted ads through tools provided by the ad networks.

But studies and interviews suggest additional considerations for protecting children, particularly young children, online. Elementary-school-age children may not understand basics of Internet safety, according to a 2005 study in the Journal of Applied Developmental Psychology. For example, such young children may not understand that submitting their name and email address for a contest can hand personal information to people who may sell it.

Such concerns are less pressing with teenagers, who tend to know as much, or more, about the Web as their parents. But a study released earlier this year by researchers at the University of California Berkeley and the University of Pennsylvania found that older teenagers are more likely than adults to believe their privacy is being protected online when it isn't. Harvard Law School professor John Palfrey, who studies children's attitudes toward privacy, said focus groups suggest that as teens learn more about behavioral targeting, they become more concerned about the practice and less likely to share information broadly.

Here are some ways that parents can help shield their children from prying eyes:

The Privacy Conversation: Parents who are concerned about behavioral tracking should first talk with their children about privacy online, says Tim Lordan, executive director of GetNetWise, a non-profit group aimed at improving online safety. Such a discussion can address many aspects of privacy on the Web—from tracking to the use of social-networking sites and the disclosure of personal information.

Parents should remind children—younger kids in particular—not to give out their name or other personal information online, Mr. Lordan said. They also should alert them to common Web tricks, such flashing pop-up ads that may install spyware on the computer. In a recent three-month period, more than 100 ad networks and exchanges delivered ads that could install malicious software on a user's computer, according to analysis by the Online Trust Association, a non-profit business-backed group.

Parents should ask elementary-school-age children to inform them when a site asks for personal information, Mr. Lordan said. The Children's Online Privacy Protection Act, known as Coppa, prohibits website operators from knowingly collecting personally identifiable information from children under 13 without parental consent. Some sites ask a child to enter a parent's email address to get approval. When parents receive a Coppa notice, they should visit the site and examine its privacy policy before allowing their child to register.

Monitoring Software: Some parents use software to monitor children's online activity, such as Net Nanny from ContentWatch Inc. and Norton Online Family from Symantec Corp. Such programs aim to prevent kids from accessing objectionable content, such as pornography, gambling or violence. They also allow parents to see and restrict social-networking activity, and some offer an option to block advertising. However, Net Nanny and Norton Online Family don't block online tracking technologies such as cookies.

Parents can use monitoring software to see when their children are visiting new sites and then inspect those sites and their privacy policies. Child-privacy experts say it's important to be candid with children about such monitoring and the reasons for it, especially because studies have shown that children see parental monitoring as an invasion of their privacy.

Blocking Ads: To prevent kids from clicking on ads, parents can consider programs that block most Web ads, and the trackers, too. Blocking ads is a controversial step because many websites rely on advertising to fund their free content. Adblock Plus, (http://adblockplus.org/en/), one widely used program, is free, though it works primarily with Mozilla's Firefox browser. After installing the program, you select filters for the ads and trackers you want to block. Two common filters, EasyList and EasyPrivacy, block most ads and trackers.

Published Sept. 17, 2010.

THE WALL STREET JOURNAL.
WSJ.com

'Scrapers' Dig Deep For Data on Web

BY JULIA ANGWIN and STEVE STECKLOW

At 1 a.m. on May 7, 2010, the website PatientsLikeMe.com noticed suspicious activity on its "Mood" discussion board. There, people exchange highly personal stories about their emotional disorders, ranging from bipolar disease to a desire to cut themselves.

It was a break-in. A new member of the site, using sophisticated software, was "scraping," or copying, every single message off PatientsLikeMe's private online forums.

PatientsLikeMe managed to block and identify the intruder: Nielsen Co., the privately held New York media-research firm. Nielsen monitors online "buzz" for clients, including major drug makers, which buy data gleaned from the Web to get insight from consumers about their products, Nielsen says.

"I felt totally violated," says Bilal Ahmed, a 33-year-old resident of Sydney, Australia, who used PatientsLikeMe to connect with other people suffering from depression. He used a pseudonym on the message boards, but his PatientsLikeMe profile linked to his blog, which contains his real name.

After PatientsLikeMe told users about the break-in, Mr. Ahmed deleted all his posts, plus a list of drugs he uses. "It was very disturbing to know that your information is being sold," he says. Nielsen says it no longer scrapes sites requiring an individual account for access, unless it has permission.

The market for personal data about Internet users is booming, and in the vanguard is the practice of "scraping." Firms offer to harvest online conversations and collect personal details from social-networking sites, resume sites and online forums where people might discuss their lives.

The emerging business of Web scraping provides some of the raw material for a rapidly expanding data economy. Marketers spent $7.8 billion on online and offline data in 2009, according to the New York management consulting firm Winterberry Group LLC. Spending on data from online sources is set to more than double, to $840 million in 2012 from $410 million in 2009.

The Wall Street Journal's examination of scraping—a trade that involves personal information as well as many other types of data—is part of the newspaper's investigation into the business of tracking people's activities online and selling details about their behavior and personal interests.

Some companies collect personal information for detailed background reports on individuals, such as email addresses, cell numbers, photographs and posts on social-network sites.

Others offer what are known as listening services, which monitor in real time hundreds or thousands of news sources, blogs and websites to see what people are saying about specific products or topics.

One such service is offered by Dow Jones & Co., publisher of the Journal. Dow Jones collects data from the Web—which may include personal information contained in news articles and blog postings—that help corporate clients monitor how they are portrayed. It says it doesn't gather information from password-protected parts of sites.

The competition for data is fierce. PatientsLikeMe also sells data about its users. PatientsLikeMe says the data it sells is anonymized, no names attached.

Nielsen spokesman Matt Anchin says the company's reports to its clients include publicly available information gleaned from the Internet, "so if someone decides to share personally identifiable information, it could be included."

Internet users often have little recourse if personally identifiable data is scraped: There is no national law requiring data companies to let people remove or change information about themselves, though some firms let users remove their profiles under certain circumstances.

California has a special protection for public officials, including politicians, sheriffs and district attorneys. It makes it easier for them to remove their home address and phone numbers from these databases, by filling out a special form stating they fear for their safety.

Data brokers long have scoured public records, such as real-estate transactions and courthouse documents, for information on individuals. Now, some are adding online information to people's profiles.

Many scrapers and data brokers argue that if information is available online, it is fair game, no matter how personal.

"Social networks are becoming the new public records," says Jim Adler, chief privacy officer of Intelius Inc., a leading paid people-search website. It offers services that include criminal background checks and "Date Check," which promises details about a prospective date for $14.95.

"This data is out there," Mr. Adler says. "If we don't bring it to the consumer's attention, someone else will."

New York-based PeekYou LLC has applied for a patent for a method that, among other things, matches people's real names to the pseudonyms they use on blogs, Twitter and other social networks. PeekYou's people-search website offers records of about 250 million people, primarily in the U.S. and Canada.

PeekYou says it also is starting to work with listening services to help them learn more about the people whose conversations they are monitoring. It says it hands over only demographic information, not names or addresses.

Employers, too, are trying to figure out how to use such data to screen job candidates. It's tricky: Employers legally can't discriminate based on gender, race and other factors they may glean from social-media profiles.

One company that screens job applicants for employers, InfoCheckUSA LLC in Florida, began offering limited social-networking data—some of it scraped—to employers about a year ago. "It's slowly starting to grow," says Chris Dugger, national account manager. He says he's particularly interested in things like whether people are "talking about how they just ripped off their last employer."

Scrapers operate in a legal gray area. Internationally, anti-scraping laws vary. In the U.S., court rulings have been contradictory. "Scraping is ubiquitous, but questionable," says Eric Goldman, a law professor at Santa Clara University. "Everyone does it, but it's not totally clear that anyone is allowed to do it without permission."

Scrapers and listening companies say what they're doing is no different from what any person does when gathering information online—they just do it on a much larger scale.

"We take an incomprehensible amount of information and make it intelligent," says Chase McMichael, chief executive of InfiniGraph, a Palo Alto, Calif., "listening service" that helps companies understand the likes and dislikes of online customers.

Scraping services range from dirt cheap to custom-built. Some outfits, such as 80Legs.com in Texas, will scrape a million Web pages for $101. One Utah company, screen-scraper.com, offers do-it-yourself scraping software for free. The top listening services can charge hundreds of thousands of dollars to monitor and analyze Web discussions.

Some scrapers-for-hire don't ask clients many questions.

"If we don't think they're going to use it for illegal purposes—they often don't tell us what they're going to use it for—generally, we'll err on the side of doing it," says Todd Wilson, owner of screen-scraper.com, a 10-person firm in Provo, Utah, that operates out of a two-room office. It is one of at least three firms in a scenic area known locally as "Happy Valley" that specialize in scraping.

Screen-scraper charges between $1,500 and $10,000 for most jobs. The company says it's often hired to conduct "business intelligence," working for companies that want to scrape competitors' websites.

One recent assignment: A major insurance company wanted to scrape the names of agents working for competitors. Why? "We don't know," says Scott Wilson, the owner's brother and vice president of sales. Another job: attempting to scrape Facebook for a multi-level marketing company that wanted email addresses of users who "like" the firm's page—as well as their friends—so they all could be pitched products.

Scraping often is a cat-and-mouse game between websites, which try to protect their data, and the scrapers, who try to outfox their defenses. Scraping itself isn't difficult: Nearly any talented computer programmer can do it. But penetrating a site's defenses can be tough.

One defense familiar to most Internet users involves "captchas," the squiggly letters that many websites require people to type to prove they're human and not a scraping robot. Scrapers sometimes fight back with software that deciphers captchas.

Some professional scrapers stage blitzkrieg raids, mounting around a dozen simultaneous attacks on a website to grab as much data as quickly as possible without being detected or crashing the site they're targeting.

Raids like these are on the rise. "Customers for whom we were regularly blocking about 1,000 to 2,000 scrapes a month are now seeing three times or in some cases 10 times as much scraping," says Marino Zini, managing director of Sentor Anti Scraping System. The company's Stockholm team blocks scrapers on behalf of website clients.

At Monster.com, the jobs website that stores resumes for tens of millions of individuals, fighting scrapers is a full-time job, "every minute of every day of every week," says Patrick Manzo, global chief privacy officer of Monster Worldwide Inc. Facebook, with its trove of personal data on some 500 million users, says it takes legal and technical steps to deter scraping.

At PatientsLikeMe, there are forums where people discuss experiences with AIDS, supranuclear palsy, depression, organ transplants, post-traumatic stress disorder and self-mutilation. These are supposed to be viewable only by members who have agreed not to scrape, and not by intruders such as Nielsen.

"It was a bad legacy practice that we don't do anymore," says Dave Hudson, who in June 2010 took over as chief executive of the Nielsen unit that scraped PatientsLikeMe in May. "It's something that we decided is not acceptable, and we stopped."

Mr. Hudson wouldn't say how often the practice occurred, and wouldn't identify its client.

The Nielsen unit that did the scraping is now part of a joint venture with McKinsey & Co. called NM Incite. It traces its roots to a Cincinnati company called Intelliseek that was founded in 1997. One of its most successful early businesses was scraping message boards to find mentions of brand names for corporate clients.

In 2001, the venture-capital arm of the Central Intelligence Agency, In-Q-Tel Inc., was among a group of investors that put $8 million into the business.

Intelliseek struggled to set boundaries in the new business of monitoring individual conversations online, says Sundar Kadayam, Intelliseek's co-founder. The firm decided it wouldn't be ethical to use automated software to log into private message boards to scrape them.

But, he says, Intelliseek occasionally would ask employees to do that kind of scraping if clients requested it. "The human being can just sign in as who they are," he says. "They don't have to be deceitful."

In 2006, Nielsen bought Intelliseek, which had revenue of more than $10 million and had just become profitable, Mr. Kadayam says. He left one year after the acquisition.

At the time, Nielsen, which provides television ratings and other media services, was looking to diversify into digital businesses. Nielsen combined Intelliseek with a New York startup it had bought called BuzzMetrics.

The new unit, Nielsen BuzzMetrics, quickly became a leader in the field of social-media monitoring. It collects data from 130 million blogs, 8,000 message boards, Twitter and social networks. It sells services such as "ThreatTracker," which alerts a company if its brand is being discussed in a negative light. Clients include more than a dozen of the biggest pharmaceutical companies, according to the company's marketing material.

Like many websites, PatientsLikeMe has software that detects unusual activity. On May 7, that software sounded an alarm about the "Mood" forum.

David Williams, the chief marketing officer, quickly determined that the "member" who had triggered the alert actually was an automated program scraping the forum. He shut down the account.

The next morning, the holder of that account e-mailed customer support to ask why the login and password weren't working. By the afternoon, PatientsLikeMe had located three other suspect accounts and shut them down. The site's investigators traced all of the accounts to Nielsen BuzzMetrics.

On May 18, PatientsLikeMe sent a cease-and-desist letter to Nielsen. Ten days later, Nielsen sent a letter agreeing to stop scraping. Nielsen says it was unable to remove the scraped data from its database, but a company spokesman later said Nielsen had found a way to quarantine the PatientsLikeMe data to prevent it from being included in its reports for clients.

PatientsLikeMe's president, Ben Heywood, disclosed the break-in to the site's 70,000 members in a blog post. He also reminded users that PatientsLikeMe also sells its data in an anonymous form, without attaching user's names to it. That sparked a lively debate on the site about the propriety of selling sensitive information. The company says most of the 350 responses to the blog post were supportive. But it says a total of 218 members quit.

In total, PatientsLikeMe estimates that the scraper obtained about 5% of the messages in the site's forums, primarily in "Mood" and "Multiple Sclerosis."

"We're a business, and the reality is that someone came in and stole from us," says PatientsLikeMe's chairman, Jamie Heywood.

Published Oct. 12, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Facebook in Privacy Breach

BY EMILY STEEL and GEOFFREY FOWLER

Many of the most popular applications, or "apps," on the social-networking site Facebook Inc. have been transmitting identifying information—in effect, providing access to people's names and, in some cases, their friends' names—to dozens of advertising and Internet tracking companies, a Wall Street Journal investigation has found.

The issue affects tens of millions of Facebook app users, including people who set their profiles to Facebook's strictest privacy settings. The practice breaks Facebook's rules, and renews questions about its ability to keep identifiable information about its users' activities secure.

The problem has ties to the growing field of companies that build detailed databases on people in order to track them online—a practice the Journal has been examining in its What They Know series. It's unclear how long the breach was in place. On Sunday, a Facebook spokesman said it is taking steps to "dramatically limit" the exposure of users' personal information.

"A Facebook user ID may be inadvertently shared by a user's Internet browser or by an application," the spokesman said. Knowledge of an ID "does not permit access to anyone's private information on Facebook," he said, adding that the company would introduce new technology to contain the problem identified by the Journal.

"Our technical systems have always been complemented by strong policy enforcement, and we will continue to rely on both to keep people in control of their information," the Facebook official said.

"Apps" are pieces of software that let Facebook's 500 million users play games or share common interests with one another. The Journal found that all of the 10 most popular apps on Facebook were transmitting users' IDs to outside companies.

The apps, ranked by research company Inside Network Inc. (based on monthly users), include Zynga Game Network Inc.'s FarmVille, with 59 million users, and Texas HoldEm Poker and FrontierVille. Three of the top 10 apps, including FarmVille, also have been transmitting personal information about a user's friends to outside companies.

Most apps aren't made by Facebook, but by independent software developers. Several apps became unavailable to Facebook users after the

Journal informed Facebook that the apps were transmitting personal information; the specific reason for their unavailability remains unclear.

The information being transmitted is one of Facebook's basic building blocks: the unique "Facebook ID" number assigned to every user on the site. Since a Facebook user ID is a public part of any Facebook profile, anyone can use an ID number to look up a person's name, using a standard Web browser, even if that person has set all of his or her Facebook information to be private. For other users, the Facebook ID reveals information they have set to share with "everyone," including age, residence, occupation and photos.

The apps reviewed by the Journal were sending Facebook ID numbers to at least 25 advertising and data firms, several of which build profiles of Internet users by tracking their online activities.

Defenders of online tracking argue that this kind of surveillance is benign because it is conducted anonymously. In this case, however, the Journal found that one data-gathering firm, RapLeaf Inc., had linked Facebook user ID information obtained from apps to its own database of Internet users, which it sells. RapLeaf also transmitted the Facebook IDs it obtained to a dozen other firms, the Journal found.

RapLeaf said that transmission was unintentional. "We didn't do it on purpose," said Joel Jewitt, vice president of business development for RapLeaf.

Facebook said it previously has "taken steps ... to significantly limit Rapleaf's ability to use any Facebook-related data."

Facebook prohibits app makers from transferring data about users to outside advertising and data companies, even if a user agrees. The Journal's findings shed light on the challenge of policing those rules for the 550,000 apps on its site.

The Journal's findings are the latest challenge for Facebook, which has been criticized in recent years for modifying its privacy rules to expose more of a user's information. This past spring, the Journal found that Facebook was transmitting the ID numbers to advertising companies, under some circumstances, when a user clicked on an ad. Facebook subsequently discontinued the practice.

"This is an even more complicated technical challenge than a similar issue we successfully addressed last spring on Facebook.com," a Facebook spokesman said, "but one that we are committed to addressing."

The privacy issue follows Facebook's effort just this month to give its users more control over its apps, which privacy activists had cited as a potential hole in users' ability to control who sees their information. On Oct. 6, Facebook created a control panel that lets users see which apps are accessing which categories of information about them. It indicates, for example, when an application accesses a user's "basic information" (including a user ID and name). However, it doesn't detail what information friends' applications have accessed about a user.

Facebook apps transform Facebook into a hub for all kinds of activity, from playing games to setting up a family tree. Apps are considered an important way for Facebook to extend the usefulness of its network. The company says 70% of users use apps each month.

Applications are also a growing source of revenue beyond advertising for Facebook itself, which sells its own virtual currency that can be used to pay for games.

Following an investigation by the Canadian Privacy Commissioner, Facebook in June 2010 limited applications to accessing only the public parts of a user's profile, unless the user grants additional permission. (Canadian officials later expressed satisfaction with Facebook's steps.) Previously, applications could tap any data the user had access to, including detailed profiles and information about a user's friends.

It's not clear if developers of many of the apps transmitting Facebook ID numbers even knew that their apps were doing so. The apps were using a common Web standard, known as a "referer," which passes on the address of the last page viewed when a user clicks on a link. On Facebook and other social-networking sites, referers can expose a user's identity.

The company says it has disabled thousands of applications at times for violating its policies. It's unclear how many, if any, of those cases involved passing user information to marketing companies.

Facebook also appeared to have shut down some applications the Journal found to be transmitting user IDs, including several created by LOLapps Media Inc., a San Francisco company backed with $4 million in venture capital. LOLapp's applications include Gift Creator, with 3.5 million monthly active users; Quiz Creator, with 1.4 million monthly active users; Colorful Butterflies and Best Friends Gifts.

A few days before this story was published, users attempting to access those applications received either an error message or were reverted to Facebook's home screen.

"We have taken immediate action to disable all applications that violate our terms," a Facebook spokesman said.

A spokeswoman for LOLapps Media declined to comment.

The applications transmitting Facebook IDs may have breached their own privacy policies, as well as industry standards, which say sites shouldn't share and advertisers shouldn't collect personally identifiable information without users' permission. Zynga, for example, says in its privacy policy that it "does not provide any Personally Identifiable Information to third-party advertising companies."

A Zynga spokeswoman said, "Zynga has a strict policy of not passing personally identifiable information to any third parties. We look forward to working with Facebook to refine how Web technologies work to keep people in control of their information."

The most expansive use of Facebook user information uncovered by the Journal involved RapLeaf. The San Francisco company compiles and sells profiles of individuals based in part on their online activities.

The Journal found that some LOLapps applications, as well as the Family Tree application, were transmitting users' Facebook ID numbers to RapLeaf. RapLeaf then linked those ID numbers to dossiers it had previously assembled on those individuals, according to RapLeaf. RapLeaf then embedded that information in an Internet-tracking file known as a "cookie."

RapLeaf says it strips out the user's name when it embeds the information in the cookie and shares that information for ad targeting. However, The Wall Street Journal found that RapLeaf transmitted Facebook user IDs to a dozen other advertising and data firms, including Google Inc.'s Invite Media.

All 12 companies said that they didn't collect, store or use the information.

Ilya Nikolayev, chief executive of Familybuilder, maker of the Family Tree application, said in an email, "It is Familybuilder's corporate policy to keep any actual, potential, current or prior business partnerships, relationships, customer details, and any similar information confidential. As this story relates to a company other than Familybuilder, we have nothing further to contribute."

Published Oct. 18, 2010.

THE WALL STREET JOURNAL.
WSJ.com

A Web Pioneer Profiles Users by Name

BY EMILY STEEL

In the weeks before the New Hampshire primary in September 2010, Linda Twombly of Nashua says she was peppered with online ads for Republican Senate hopeful Jim Bender.

It was no accident. An online tracking company called RapLeaf Inc. had correctly identified her as a conservative who is interested in Republican politics, has an interest in the Bible and contributes to political and environmental causes. Mrs. Twombly's profile is part of RapLeaf's rich trove of data, garnered from a variety of sources and which both political parties have tapped.

RapLeaf knows even more about Mrs. Twombly and millions of other Americans: their real names and email addresses.

This makes RapLeaf a rare breed. Rival tracking companies also gather minute detail on individual Americans: They know a tremendous amount about what you do. But most trackers either can't or won't keep the ultimate piece of personal information—your name—in their databases. The industry often cites this layer of anonymity as a reason online tracking shouldn't be considered intrusive.

RapLeaf says it never discloses people's names to clients for online advertising. But possessing real names means RapLeaf can build extraordinarily intimate databases on people by tapping voter-registration files, shopping histories, social-networking activities and real estate records, among other things.

"Holy smokes," says Mrs. Twombly, 67 years old, after The Wall Street Journal decoded the information in RapLeaf's file on her. "It is like a watchdog is watching me, and it is not good."

Some early adopters of the service are political campaigns. Democratic political consultant Chris Lehane used RapLeaf in a successful campaign against Proposition 17 in California, which would have changed the way auto-insurance rates are set in the state.

RapLeaf ranks among the most sophisticated players in the fast-growing business of profiling people online and trading in personal details of their lives, an

industry that is the focus of a Journal investigation. The San Francisco startup says it has 1 billion email addresses in its database.

RapLeaf acknowledges collecting names. It says it doesn't include Web-browsing behavior in its database, and it strips out names, email addresses and other personally identifiable data from profiles before selling them for online advertising.

Nevertheless, the Journal found that, in certain circumstances, RapLeaf had transmitted identifying details about Mrs. Twombly—such as a unique Facebook ID number, which can be linked back to a person's real name—to at least 12 companies. The Journal also found RapLeaf had transmitted a unique MySpace ID number (which is sometimes linked to a person's real name), to six companies. MySpace is owned by News Corp., which publishes the Journal.

RapLeaf says its transmission of Facebook and MySpace IDs was inadvertent and the practice was ended after the Journal brought it to the company's attention. The company says people can permanently opt out of its services at RapLeaf.com.

RapLeaf executives say their business offers valuable consumer benefits by allowing people to see relevant advertising and content. "The key goal of RapLeaf is to build a more personalizable world for people," says RapLeaf CEO Auren Hoffman. "We think a more personalizable world is a better world."

When a person logs in to certain sites, the sites send identifying information to RapLeaf, which looks up that person in its database of email addresses.

Then, RapLeaf installs a "cookie," a small text file, on the person's computer containing details about the individual (minus name and other identifiable facts). Sites where this happened include e-card provider Pingg.com and advice portal About.com and picture service TwitPic.com.

In some cases, RapLeaf also transmits data about the person to advertising companies it partners with.

Data gathered and sold by RapLeaf can be very specific. According to documents reviewed by the Journal, RapLeaf's segments recently included a person's household income range, age range, political leaning, and gender and age of children in the household, as well as interests in topics including religion, the Bible, gambling, tobacco, adult entertainment and "get rich quick" offers.

In all, RapLeaf segmented people into more than 400 categories, the documents indicated.

RapLeaf's privacy policy states it won't "collect or work with sensitive data on children, health or medical conditions, sexual preferences, financial account information or religious beliefs."

After the Journal asked RapLeaf whether some of its profile segments contradicted its privacy policy, the company eliminated many of those segments. Segments eliminated include: interest in the Bible, Hispanic and Asian ethnic

products, gambling, tobacco, adult entertainment, "get rich quick" offers and age and gender of children in household.

RapLeaf says many of its segments are also "used widely by the direct-marketing industry today."

In 2010's hotly contested midterm elections, some political organizations are tapping RapLeaf's technology. With traditional postal mailing lists, "We used to bombard their house with mail. Now we can bombard their house with online ads," says Robert Willington, the Republican online campaign strategist who worked on behalf of Mr. Bender's New Hampshire campaign.

RapLeaf helped Mr. Bender's campaign target likely Republican voters with ads online. (Mr. Bender, who confirms working with RapLeaf, lost the election.)

In Mr. Lehane's California effort against Proposition 17 in 2010, RapLeaf found online about 200,000 suburban women over the age of 40 in Southern California, a demographic the campaign considered swing voters.

Mr. Lehane says the 4-percentage-point margin of defeat suggested the technology was effective. "With an election that close, every voter you can reach matters," he says.

Mr. Lehane says he was considering using RapLeaf as part of a campaign against Meg Whitman, who is running for governor in California. That campaign is being run by a political group, Level the Playing Field 2010, which was funded by several labor unions and which Mr. Lehane led.

RapLeaf says it has participated in about 10 campaigns during the 2010 season, declining to identify them. "We expect that forward-thinking campaigns will begin to use it this year more widely as an alternative to direct mail, email and phone calls," says Joel Jewitt, RapLeaf's vice president of business development.

Co-founded in 2006 by Mr. Hoffman, a Silicon Valley entrepreneur, RapLeaf began as an online service letting people rate each other based on their business transactions.

The company raised an initial $1 million in funding from well-known Silicon Valley investors including PayPal co-founder and Facebook investor Peter Thiel. A person familiar with the situation says the company closed a $15 million fund-raising round in October 2010.

Soon after it was founded, RapLeaf began "scraping"—or collecting information from—social networks to build a people search engine. It matched data from social-networking profiles with email addresses. RapLeaf says data it collects are public. It sold a service giving companies information about the customers on their e-mail lists.

By 2009, RapLeaf had indexed more than 600 million unique email addresses, it said in a press release that year, and was adding more at a rate of 35 million per month. Meanwhile, the business of helping marketers with their

email lists (RapLeaf's core) was lagging in the recession. And the online-tracking business was taking off.

RapLeaf's Mr. Jewitt says the company saw an opportunity: It decided to connect its database of dossiers on people to cookies placed on those same individuals' computers, for ad targeting. "If you are a modern information company, you have to be involved in that," he says.

Combining off-line profiles with online tracking has raised red flags ever since another company first tried it 10 years ago. Privacy advocates argued that connecting people's Web-browsing habits with their names was too intrusive.

RapLeaf says it doesn't share or sell emails. However, under some circumstances it will provide names and other personal details if a client already possesses that person's email address.

For example, a company might come to RapLeaf with an email-address mailing list, and RapLeaf will try to provide information about the people on that list. This year, RapLeaf began offering services to target these people with online ads for the client.

For that to work, RapLeaf relies on a network of cooperating websites that use email addresses as part of the sign-on process. Those sites agree to transmit their users' email addresses (in encrypted form) to RapLeaf. Then, RapLeaf "drops," or installs, cookies on users' computers.

It's tough to build up a network of such sites, because many don't want to let outsiders track their visitors. In summer 2010, RapLeaf sent a marketing email offering to pay one website an unspecified sum for this kind of access, according to documents reviewed by the Journal. The website chose not to take the offer.

RapLeaf declined to name the sites it works with, citing nondisclosure agreements. The Journal found that sites installing RapLeaf cookies included About.com, owned by the New York Times Co.; online invitation site Pingg.com; photo-sharing sites TwitPic.com and Plixi.com; movie site Flixster.com; discount site Tester-Rewards.com; and some Facebook.com and MySpace.com applications.

The Journal previously reported on the Facebook and MySpace apps sending data to RapLeaf. Both sites say they prohibit applications from sharing user data with outside data companies, and that they took steps to stop the apps that were transmitting user data to RapLeaf.

A Facebook spokesman says the company is acting to "dramatically limit" the exposure of users' personal information. Facebook says the user ID allows access only to information that Facebook requires people to make public in their profile.

MySpace says it uses RapLeaf data for its "friend recommendation" system, but doesn't share user data or let RapLeaf track MySpace users.

After receiving user IDs from some MySpace and Facebook apps, RapLeaf was then transmitting data about users to its advertising partners. After

being contacted by the Journal, RapLeaf says it "acted immediately" to strip out identifying information from the data it shared with partners.

An About.com spokeswoman says the company doesn't have a relationship with RapLeaf. She says users' information was sent to RapLeaf via a partner that operates on its site, and that About.com wasn't aware its users' email addresses were being sent to RapLeaf.

Plixi.com says the company is "in experiment mode right now with behavioral-targeting companies like RapLeaf." Flixster.com says it "does not sell any of our users' personal information to anyone" and declined to comment further.

Pingg.com declined to comment. TwitPic and Tester-Rewards didn't respond to requests for comment.

The Journal decoded RapLeaf's information on Gordon McCormack Jr., a 52-year-old who lives in Ashland, N.H. RapLeaf correctly identified Mr. McCormack's income range, number of cars (one), his interests in gardening and the Beatles, and his interest in playing the online game Mafia Wars, among other topics.

Mr. McCormack says he plays Mafia Wars almost every day before going to bed.

RapLeaf also identified Mr. McCormack as someone with an interest in online personals. He says he isn't currently active in online dating, but might have a couple of profiles "lurking on the Internet."

When Mrs. Twombly, the New Hampshire Republican, registered at Pingg.com using her email address, RapLeaf matched her to dozens of "segments," according to a Journal analysis of the computer code transmitted while she was on the site.

The Journal was able to decode 26 of the segments, including her income range and age range and the fact that she is interested in the Bible and in cooking, crafts, rural farming and wildlife. Mrs. Twombly says all the decoded segments describe her accurately.

RapLeaf says some of the segments in Mrs. Twombly's and Mr. McCormack's profiles "do not exist," possibly due to changes in RapLeaf's overall segment list in the time since their Web traffic was decoded for this article in September 2010.

In Mrs. Twombly's case, RapLeaf transmitted data about her to at least 23 data and advertising companies after she logged into Pingg, according to the analysis of the computer code.

Twenty-two companies, including Google's Invite Media, confirmed receiving data from RapLeaf. RapLeaf declined to comment on its relationships with the companies.

Since talking with the Journal, Mrs. Twombly tweaked her Web browser to limit cookie installation. As a result, she says, some websites don't always work properly for her, a common side effect of restricting cookies.

Mrs. Twombly also removed applications from her Facebook profile that were transmitting data to RapLeaf, the Best Friends Gifts and Colorful Butterflies apps. The maker of those apps, LOLapps Media Inc., says it stopped working with RapLeaf.

Still, Mrs. Twombly is no longer using those apps to send virtual gifts and butterflies to her online friends. "My neighbor did send me a hug or a rainbow or a heart or something like that, but I didn't respond," Mrs. Twombly says. "Once burned, twice shy."

Julia Angwin and Peter Wallsten contributed to this article.

Published Oct. 25, 2010.

For interactive graphics related to this story, click this link.

THE WALL STREET JOURNAL.
WSJ.com

Politicians Tap Sophisticated Online Tracking Tools

BY EMILY STEEL and PETER WALLSTEN

Politicians are deploying sophisticated new technologies to track Internet users—sometimes by name—and identify their political leanings.

Officials and consultants with both major parties are using the techniques ahead of next month's election, which include matching voter names from registration rolls with online profiles and studying voters' online "body language."

"As political professionals, the more data we have, the happier we are," said Kristen Luidhardt, a Republican consultant in Indianapolis. "We'd love to know absolutely everything about you."

Technology can be key in close elections. Republicans in 2004 used their Voter Vault database—which combined voter registrations with consumer marketing data—to help George W. Bush win reelection. Four years later, Barack Obama used social-networking tools.

One technique matches voters' names and addresses to their online behavior. RapLeaf Inc. and AOL Inc. offer services to show ads targeted to specific groups of voters. Consultants from both parties have used RapLeaf technology in about 10 races this year nationwide.

In AOL's case, the company matches voters with people who use AOL services, such as email or instant messaging. Information is then added to a cookie on the user's computer, allowing the campaign to target that person with ads.

AOL says an outside company does the matching. It says information in its cookies can't be tracked back to an individual, and that people can opt out by visiting its website or deleting their cookies.

"This is one of the best targeting options out there," says Andrew Bleeker, the Democratic consultant who directed Internet advertising for Mr. Obama's presidential campaign. Mr. Bleeker says he is considering using online-offline matching services to target voters in the final days of the campaign. He declines to identify his clients.

Privacy advocates argue that connecting people's Web-browsing habits with information tied to their names is too intrusive.

The Republican National Committee in the summer of 2010 built a data system it calls the "Blender" to target voters for fund-raising and other appeals. When people register for a campaign activity or give money through the RNC website, the Blender attaches voter records and other information about the user. The RNC then sends those people targeted messages such as emails.

One company behind the RNC's technology is Eloqua Ltd., of Vienna, Va. Chief Executive Joe Payne says Eloqua analyzes people's online reading habits to understand a user's "digital body language." When the user enters a name or home address on a website using Eloqua technology, this previously anonymous information can be matched to his name.

An official of the Democratic National Committee says the Democrats also merge online and offline data for "customizing, localizing and personalizing" voter interactions.

Tommy Sowers, a Missouri Democrat seeking a U.S. House seat, is tracking people who visit his campaign site, and targeting them with ads on other sites, according to his online consultant.

Both parties say voter privacy is a primary concern.

Targeted Victory, a Republican consulting firm, says it is targeting voters through technology designed by the online ad firm Lotame Inc. Lotame can analyze the public comments people post on websites to determine voter "sentiment." The company says the commenters it monitors remain anonymous.

During the summer, Lotame said it analyzed comments on HuffingtonPost.com for several weeks, and then stopped.

A Huffington Post spokesman says Lotame's actions weren't authorized and came "as a complete surprise." Huffington Post currently uses Lotame services to place ads on its site.

Lotame says it analyzes comments only when it has an agreement with the publisher.

Republican consultant CampaignGrid LLC is using Internet targeting another way—to beam political ads to very small geographic areas. Founder Richard Masterson says the firm can target voters in a ZIP+4 code, which can include just a few homes.

Nevada Senate Republican candidate Sharron Angle used CampaignGrid to target ads to Republican areas not just in Nevada, but also in Southern California, Texas and elsewhere, according to CampaignGrid. A campaign spokesman declined to comment.

Mr. Masterson says the targeting usually isn't that precise, and that a dozen homes is more typical. He says the targeting is based purely on location and not personal characteristics.

Published Oct. 25, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Insurers Test Data Profiles To Identify Risky Clients

BY LESLIE SCISM and MARK MAREMONT

Life insurers are testing an intensely personal new use for the vast dossiers of data being amassed about Americans: predicting people's longevity.

Insurers have long used blood and urine tests to assess people's health—a costly process. Today, however, data-gathering companies have such extensive files on most U.S. consumers—online shopping details, catalog purchases, magazine subscriptions, leisure activities and information from social-networking sites—that some insurers are exploring whether data can reveal nearly as much about a person as a lab analysis of their bodily fluids.

In one of the biggest tests, the U.S. arm of British insurer Aviva PLC looked at 60,000 recent insurance applicants. It found that a new, "predictive modeling" system, based partly on consumer-marketing data, was "persuasive" in its ability to mimic traditional techniques.

The research heralds a remarkable expansion of the use of consumer-marketing data, which is traditionally used for advertising purposes.

This data increasingly is gathered online, often with consumers only vaguely aware that separate bits of information about them are being collected and collated in ways that can be surprisingly revealing.

The growing trade in personal information is the subject of a Wall Street Journal investigation into online privacy.

A key part of the Aviva test, run by Deloitte Consulting LLP, was estimating a person's risk for illnesses such as high blood pressure and depression. Deloitte's models assume that many diseases relate to lifestyle factors such as exercise habits and fast-food diets.

This kind of analysis, proponents argue, could lower insurance costs and eliminate an off-putting aspect of the insurance sale for some people.

"Requiring every customer to provide additional, and often unnecessary, information" such as blood or urine samples, "simply makes the process less efficient and less customer-friendly," says John Currier, chief actuary for Aviva USA.

Other insurers exploring similar technology include American International Group Inc. and Prudential Financial Inc., executives for those firms confirm.

Deloitte, a big backer of the concept, has pitched it in recent months to numerous insurers.

The industry is grappling with how to get policies into the hands of middle-class families more cost-effectively. Sales of life policies to individuals are down 45% since the mid-1980s. Deloitte says insurers could save $125 per applicant by eliminating many conventional medical requirements.

Under Deloitte's predictive model, the cost to achieve similar results would be $5, Deloitte says. The total underwriting costs for a policy range from $250 to $1,000, insurers say.

Making the approach feasible is a trove of new information being assembled by giant data-collection firms. These companies sort details of online and offline purchases to help categorize people as runners or hikers, dieters or couch potatoes.

They scoop up public records such as hunting permits, boat registrations and property transfers. They run surveys designed to coax people to describe their lifestyles and health conditions.

Increasingly, some gather online information, including from social-networking sites. Acxiom Corp., one of the biggest data firms, says it acquires a limited amount of "public" information from social-networking sites, helping "our clients to identify active social-media users, their favorite networks, how socially active they are versus the norm, and on what kind of fan pages they participate."

For insurers and data-sellers alike, the new techniques could open up a regulatory can of worms. The information sold by marketing-database firms is lightly regulated.

But using it in the life-insurance application process would "raise questions" about whether the data would be subject to the federal Fair Credit Reporting Act, says Rebecca Kuehn of the Federal Trade Commission's division of privacy and identity protection.

The law's provisions kick in when "adverse action" is taken against a person, such as a decision to deny insurance or increase rates.

The law requires that people be notified of any adverse action and be allowed to dispute the accuracy or completeness of data, according to the FTC.

Deloitte and the life insurers stress the databases wouldn't be used to make final decisions about applicants. Rather, the process would simply speed up applications from people who look like good risks. Other people would go through the traditional assessment process.

The use of the data also may require passing muster with insurance regulators. Regulators in Connecticut, New Jersey and New York, all home to major U.S. life insurers, say they haven't been briefed.

They say their concerns would include ensuring that the approach doesn't unfairly discriminate. "An insurer could contend that a subscription to 'Hang Gliding Monthly' is predictive of highly dangerous behavior, but I'm not buying

that theory: The consumer may be getting the magazine for the pictures," says Thomas Considine, New Jersey's commissioner of banking and insurance.

AIG is in the early stages of analysis "to figure out what is meaningful and what is not" in the data, says Bob Beuerlein, chief actuary for its SunAmerica Financial unit. The tests are being conducted by an in-house "think tank" whose mission, he says, is "to see where we're going in the future."

A Prudential spokesman says the insurer "is looking at" the potential of marketing data, declining to discuss details.

Some insurers are taking a wait-and-see approach. Deloitte's "methodology is sound," says Mike Belko, chief underwriter at USAA Life Insurance Co., but for now, "it's too soon to say how much reliance we would put on the information."

The largest marketing-database companies in the U.S. include Acxiom, Alliance Data Systems Corp., Experian PLC, and Infogroup. Each says it has detailed information on more than 100 million U.S. households, though contents of their databases vary as do their rules related to data use.

There are myriad sources of personal data. Acxiom recently told investors it takes in three billion pieces of information daily as businesses seek to "monetize" information about their customers. Some retailers share information about purchases made by people, including item description, price and the person's name.

Increasingly, information comes from people's online behavior. Acxiom says it buys data from online publishers about what kinds of articles a subscriber reads—financial or sports, for example—and can find out if somebody's a gourmet-food lover from their online purchases. Online marketers often tap data sources like these to target ads at Web users.

"Personally identifiable data from the online world is merged with personally identifiable information from the offline world, every day," says Jennifer Barrett, Acxiom's head of global privacy and public policy. She also says that, while Acxiom does store personally identifiable information, it doesn't store or merge anonymous online-tracking data, such as Web-browsing records.

Acxiom says it wouldn't let insurers use its data to help assess applicants, for fear of triggering the stiffer federal credit-reporting regulations. Infogroup says it isn't supplying information to insurers for this use. Experian said its marketing data may only be used for marketing purposes.

Units of News Corp., including The Wall Street Journal, supply information to marketing-database firms and buy information from them. "We have strict precautions around confidentiality," a spokeswoman said.

This isn't the first use of database mining in insurance. About 20 years ago, data pros found that some factors in people's credit histories have a strong correlation to claims on car and home-insurance policies.

In other words: The better your credit, the less likely you'll file a claim. Today, most car and home insurers use this phenomenon to price their policies. For this purpose, property-casualty insurers look at people's credit reports, as opposed to the consumer-marketing databases.

Life insurers haven't changed their general underwriting approach for decades, relying heavily on medical screening.

Deloitte's effort to promote predictive modeling to life insurers gained steam in recent months, boosted partly by the Aviva research. Deloitte detailed the test in May 2010 at a seminar hosted by the Society of Actuaries, a professional group.

At the seminar, a consultant helped explain Deloitte's concept by discussing imaginary 40-year-old insurance buyers, "Beth" and "Sarah."

Using readily available data, the consultant said, an insurer could learn that Beth commutes some 45 miles to work, frequently buys fast food, walks for exercise, watches a lot of television, buys weight-loss equipment and has "foreclosure/bankruptcy indicators," according to slides used in the presentation.

"Sarah," on the other hand, commutes just a mile to work, runs, bikes, plays tennis and does aerobics. She eats healthy food, watches little TV and travels abroad. She is an "urban single" with a premium bank card and "good financial indicators."

Deloitte's approach, the consultant said, indicates Sarah appears to fall into a healthier risk category.

Beth seems to be a candidate for a group with worse-than-average predicted mortality. The top five reasons: "Long commute. Poor financial indicators. Purchases tied to obesity indicators. Lack of exercise. High television consumption indicators."

Another consultant detailed the Aviva test to the seminar attendees. Deloitte didn't identify the insurer; Aviva confirmed its role to the Journal.

The consumer-marketing data for the test came from Equifax Inc.'s marketing-services unit, since bought by Alliance Data Systems. An Alliance spokeswoman says the company was unaware of the insurance-related test, which was done before it bought the former Equifax subsidiary. Alliance "does not provide its marketing data for such purposes," she says.

The goal of Aviva's test: With 60,000 actual insurance applicants, figure out how to use the marketing databases and other information to reach the same underwriting conclusions that Aviva reached using traditional methods such as blood work. The 60,000 people were applicants Aviva had already judged.

Such predictive models wouldn't necessarily look for indicators of all diseases, such as AIDS, because the insurer would likely learn about some conditions from the answers on an application. Rather, insurers say a model would tend to look for potential risks such as, for instance, diabetes (from, say, a poor diet).

Aviva declined to discuss the process in detail, but Mr. Currier says the insurer found that the model consistently yielded results that "closely aligned with those of purely traditional underwriting decisions."

The insurer says pilot projects with marketing data are continuing in its effort to improve clients' buying experience.

Deloitte acknowledges the potentially controversial nature of its work. "No matter what their predictive powers may be, any variable that is deemed to create a legal or public-relations risk, or is counter to the company's 'values,' should be excluded from the model," its consultants wrote in an April 2010 paper.

Deloitte isn't the only firm pushing data-mining for insurers. Celent, an insurance consulting arm of Marsh & McLennan Cos., recently published a study suggesting insurers could use social-networking data to help price policies and aid in fraud detection.

A life insurer might want to scrutinize an applicant who reports no family history of cancer, but indicates online an affinity with a cancer-research group, says Mike Fitzgerald, a Celent senior analyst.

"Whether people actually realize it or not, they are significantly increasing their personal transparency," he says. "It's all public, and it's electronically mineable."

Published Nov. 19, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Inside Deloitte's Life-Insurance Assessment Technology

Deloitte Consulting LLP's effort to persuade life insurers that marketing data can size up people's longevity faces an obstacle: How to know if the method works, when the proof—policyholder deaths—is years away?

So the company developed a test involving 60,000 insurance applicants at the U.S. arm of British insurer Aviva PLC. Deloitte detailed the findings this spring at an industry seminar.

Here's how the test worked. Aviva used traditional underwriting methods—including costly blood and urine tests—to assess the 60,000 applicants. Aviva sorted those people into various risk categories. In addition, some of the individuals had been declined a policy.

Deloitte then took the 60,000 cases and tried to replicate Aviva's traditional underwriting decisions with a new methodology not reliant on blood work.

First, it divided the 60,000 into two equal groups. (Names were withheld.) For the first 30,000, it studied the traditional underwriting decisions and set out to build a data-only "predictive model" that could reach similar conclusions.

Its model didn't use the blood and urine tests. It would be designed to use data such as personal and family medical histories, as detailed on written insurance-application forms. Deloitte's team also had access to industry-shared information from past insurance applications and motor-vehicle reports. On top of that, it used the consumer-marketing data.

The consumer data came from Equifax Inc.'s marketing-services unit, since acquired by Alliance Data Systems Corp. The data noted which of hundreds of attributes applied to each individual—likely hobbies, TV-viewing habits, income estimates. In Deloitte's final predictive model, the nontraditional consumer-marketing data represented about 37% of the predictive ability, it says.

Then came the real test: Running the predictive model on the second batch of 30,000 applicants, to see if it could accurately replicate the underwriters' original assessments. Aviva and Deloitte judged the test largely successful. "The use of third-party data was persuasive across the board in all cases," said John Currier, chief actuary for Aviva USA.

Published Nov. 19, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Shunned Profiling Method On the Verge of Comeback

BY STEVE STECKLOW and PAUL SONNE

One of the most potentially intrusive technologies for profiling and targeting Internet users with ads is on the verge of a comeback, two years after an outcry by privacy advocates in the U.S. and Britain appeared to kill it.

The technology, known as "deep packet inspection," is capable of reading and analyzing the "packets" of data traveling across the Internet. It can be far more powerful than "cookies" and other techniques commonly used to track people online because it can be used to monitor all online activity, not just Web browsing. Spy agencies use the technology for surveillance.

Now, two U.S. companies, Kindsight Inc. and Phorm Inc., are pitching deep packet inspection services as a way for Internet service providers to claim a share of the lucrative online ad market.

Kindsight and Phorm say they protect people's privacy with steps that include obtaining their consent. They also say they don't use the full power of the technology, and refrain from reading email and analyzing sensitive online activities.

Use of deep packet inspection this way would nonetheless give advertisers the ability to show ads to people based on extremely detailed profiles of their Internet activity. To persuade Internet users to opt in to be profiled, Kindsight will offer a free security service, while Phorm promises to provide customized Web content such as news articles tailored to users' interests. Both would share ad revenue with the ISPs.

Kindsight says its technology is sensitive enough to detect whether a particular person is online for work, or for fun, and can target ads accordingly.

"If you're trying to engage in one-stop-shopping surveillance on the Internet, deep packet inspection would be an awesome tool," says David C. Vladeck, director of the Federal Trade Commission's Bureau of Consumer Protection. When deep packet inspection is used for targeted ads, the FTC has made it clear that broadband providers "should, at a minimum, notify consumers that the ISP was mining the information and obtain clear consumer consent," Mr. Vladeck says.

Kindsight, majority owned by telecommunications giant Alcatel-Lucent SA, says six ISPs in the U.S., Canada and Europe have been testing its security service this year although it isn't yet delivering targeted ads. It declined to name the clients.

"These are tier-one ISPs we're working with," says Mike Gassewitz, Kindsight's chief executive. He says his company also has been placing ads on various websites to test the ad-placement technology and build up a base of advertisers, which now number about 100,000.

Two large ISPs in Brazil—Oi, a unit of Tele Norte Leste Participacoes SA, and Telefonica SA—currently have deals with Phorm. Oi, Brazil's largest broadband provider with about 4.5 million customers, has launched the product initially with about 10,000 people in Rio De Janeiro.

"We want to grow that," says Pedro Ripper, Oi's strategy and technology director.

A spokesman for Telefonica says it is testing the service on about 1,000 broadband customers and will evaluate the results before deciding whether to roll it out. "The user has the choice to enable or disable the service anytime he or she wants to," the company said in a statement.

Phorm is hoping to introduce its service in South Korea and eventually in the U.S. "It is designed from the ground up to ensure one thing and that is privacy," says Kent Ertugrul, Phorm's chief executive.

Kindsight and Phorm say the ISPs don't provide them with subscribers' real identities. Both also say they don't collect any personal information, read email, store users' browsing histories or monitor sensitive sites such as health blogs. Subscribers must "opt in," or give their consent to participate, both companies say.

Both the Kindsight and Phorm systems study people's behavior and interests based on the websites they visit to show them relevant ads. Mr. Gassewitz says that unlike Web-based tracking methods, which generally create a single behavioral profile no matter how many people share a computer, Kindsight can "generate multiple characters per human."

"If I come online and I'm in work mode, I will show up as a very different character than when I go online Saturday morning and I'm in recreation mode," he says. The targeted ads would reflect which "character" is online.

Mr. Gassewitz calls that some of Kindsight's "secret sauce." The company this year filed a patent on its "character differentiation" technology.

A new revenue source would mark a welcome change for ISPs. The companies have been under pressure to offer ever-faster Internet services at lower prices, while Google Inc. and other companies raked in billions of dollars selling ads. Targeted ads based on people's interests or behavior generally fetch higher fees.

ISPs "feel like they have data and they ought to be able to use it," says Tim McElgunn, chief analyst at Pike & Fischer Broadband Advisory Services. "They really desperately want to."

This isn't the first time ISPs have tried this. Two years ago, ISPs in the U.S. and Britain signed deals with companies offering deep packet inspection services and a cut of ad revenue.

Those pacts fell apart after a privacy outcry. In the U.K., an uproar ensued after BT Group PLC admitted it had tested Phorm's technology on some subscribers without telling them. In 2009, BT and two other British ISPs that explored deploying Phorm's service—Virgin Media Inc. and TalkTalk—abandoned it.

In the U.S., controversy erupted in 2008 over the practices of a company called NebuAd Inc., which planned to use deep packet inspection to deliver targeted advertising to millions of broadband subscribers unless they explicitly opted out of the service. At a congressional hearing, Bob Dykes, the company's founder, was grilled over its policy. NebuAd stopped doing business in 2009; several U.S. ISPs who signed deals with NebuAd have been hit with class-action lawsuits accusing them of "installing spyware devices" on their networks.

In an interview, Mr. Dykes said, "If I had to do things over again, I would have figured out how to architect an opt-in model."

The companies now offering ad services based on deep packet inspection believe they have learned how to make the services acceptable to privacy advocates and Internet users. This includes asking for permission up front and offering people incentives to receive targeted ads, such as Kindsight's free security service, which includes identity-theft protection. Customers can pay a monthly fee to receive no ads.

In Brazil, Phorm is emphasizing customized content on partner websites if people agree to opt in. For example, users visiting a sports website might see articles about their favorite teams (gleaned from an analysis of their surfing habits), providing an online experience different from other people.

"Receive your favorite content in an easy and practical way and without spending money!" says Oi's main opt-in screen for the Phorm service, called Navegador. "We guarantee your privacy! No personal information is input in the program, so your privacy is guaranteed!"

Oi's Mr. Ripper says more than half the subscribers offered the service in the initial launch have opted in to date. "We were very happy with it," he says. He says two outside auditors verified Phorm's privacy-protection settings.

Until 2007, Phorm was known as 121Media Inc. It delivered targeted ads, particularly pop-ups, to users who downloaded free software. The ads were "based on an anonymous analysis of their browsing behavior, which is likely to indicate their commercial and lifestyle interests," according to corporate filings.

Several Internet security companies, including Symantec Corp., flagged part of 121Media's adware system as "spyware." Microsoft's Malware Protection Center called it a "trojan," or malicious software disguised as something useful.

Facing "a combination of public perception and legal and technological challenges," 121Media said it shifted its focus in 2005 from the desktop-adware business to ISPs.

It eventually shuttered its adware business and renamed itself Phorm. The company is led by Mr. Ertugrul, a Princeton-educated, former investment banker who in the early 1990s formed a joint venture with the Russian Space Agency to offer joy rides to tourists in MiG-29 fighter jets. The venture was later sold.

In February 2008, Britain's biggest ISPs—BT, Virgin Media and TalkTalk—announced plans to implement Phorm's service. Those plans quickly unraveled.

Suspicions earlier had arisen among some BT subscribers who discovered they were being routed through an unfamiliar Internet address when they tried to visit a website. Some of them contacted BT and were advised their computer might be infected with a virus, according to a person familiar with the matter.

A BT spokesman said it is "standard procedure" to take customers through "a number of steps to try and identify the issue" if they call with a question about their service.

In fact, the subscribers were part of tests BT conducted in 2006 and 2007 using Phorm's technology. When BT disclosed the testing in April 2008, the backlash was fierce, with online protests by privacy advocates and government investigations. Four members of the board of directors later resigned, including former AT&T chief executive David Dorman and ex-Coca-Cola Co. president Steven Heyer, citing differences with Mr. Ertugrul. Messrs. Dorman and Heyer declined to comment.

The three ISPs eventually bailed out. "Phorm was bad news," says David Smith, deputy commissioner of Britain's Information Commissioner's Office, which oversees data protection. He says he's not surprised Phorm is looking for clients abroad. "It was pretty clear that no one was going to touch them in the UK."

Kindsight's roots trace to an in-house project known as Project Rialto at Alcatel-Lucent, where Mr. Gassewitz once worked as a vice president of strategic planning.

A 2007 job posting on Project Rialto's website described the company's work as developing "systems that can handle [a] massive volume of data for in-depth analysis of user behavior to enable targeted advertising."

Project Rialto eventually became Kindsight, a spinoff. At an Alcatel-Lucent conference held in September 2008 in Beverly Hills, Mr. Gassewitz spoke at a session called "Merging Technology and Advertising." A summary of his comments, posted on Alcatel's website, reads in part: "Through technologies like

deep packet inspection," Internet service providers "can gather even more information about consumers" than rivals such as Google or Facebook.

Mr. Gassewitz also talked about "significant privacy concerns," the summary says, and stressed that ISPs must find a way to provide measurable value to consumers "to avoid backlash."

To win over Internet users to its services, Kindsight plans to offer what it has described as a "free, always-on, always-up-to-date security service."

"Say hello to your new best friend. . ." it said on its redesigned website in 2008. The company later dropped the slogan. "That was early days," says Mr. Gassewitz.

Before giving away the security service free, Kindsight plans to display an opt-in screen to ISP users that explains how its technology analyzes "websites visited and searches conducted to assign a numerical value to various interest categories." The "score" is used to deliver relevant ads.

In market-research tests in North America, France and the U.K., Kindsight found that about 60% of users were willing to take the service free in exchange for receiving targeted ads, he says. Another 10% were willing to pay for it.

Mr. Gassewitz says six ISPs have tested Kindsight's security service on subscriber groups as big as 200,000. Mr. Gassewitz says, "There was no profiling occurring, no advertising occurring, no data collection occurring."

Oi's Mr. Ripper believes that the technology's time has come. "The Internet is becoming more and more a platform to deliver very targeted messages," he says. As for deep packet inspection, "Everyone is going to get there. It's just a matter of timing."

Published Nov. 24, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Race Is On To 'Fingerprint' Phones, PCs

BY JULIA ANGWIN and JENNIFER VALENTINO-DEVRIES

IRVINE, Calif.—David Norris wants to collect the digital equivalent of fingerprints from every computer, cellphone and TV set-top box in the world.

He's off to a good start. So far, Mr. Norris's start-up company, BlueCava Inc., has identified 200 million devices. By the end of next year, BlueCava says it expects to have cataloged one billion of the world's estimated 10 billion devices.

Advertisers no longer want to just buy ads. They want to buy access to specific people. So, Mr. Norris is building a "credit bureau for devices" in which every computer or cellphone will have a "reputation" based on its user's online behavior, shopping habits and demographics. He plans to sell this information to advertisers willing to pay top dollar for granular data about people's interests and activities.

Device fingerprinting is a powerful emerging tool in this trade. It's "the next generation of online advertising," Mr. Norris says.

It might seem that one computer is pretty much like any other. Far from it: Each has a different clock setting, different fonts, different software and many other characteristics that make it unique. Every time a typical computer goes online, it broadcasts hundreds of such details as a calling card to other computers it communicates with. Tracking companies can use this data to uniquely identify computers, cellphones and other devices, and then build profiles of the people who use them.

Until recently, fingerprinting was used mainly to prevent illegal copying of computer software or to thwart credit-card fraud. BlueCava's own fingerprinting technology traces its unlikely roots to an inventor who, in the early 1990s, wanted to protect the software he used to program music keyboards for the Australian pop band INXS.

Tracking companies are now embracing fingerprinting partly because it is much tougher to block than other common tools used to monitor people online, such as browser "cookies," tiny text files on a computer that can be deleted.

As controversy grows over intrusive online tracking, regulators are looking to rein it in. This week, the Federal Trade Commission is expected to release a privacy report calling for a "do-not-track" tool for Web browsers.

Ad companies are constantly looking for new techniques to heighten their surveillance of Internet users.

Deep packet inspection, a potentially intrusive method for peering closely into the digital traffic that moves between people's computers and the broader Internet, is being tested in the U.S. and Brazil as a future means to deliver targeted advertising.

Akamai Technologies Inc., an Internet-infrastructure giant that says it delivers 15% to 30% of all Web traffic, is marketing a technique to track people's online movements in more detail than traditional tools easily can.

It's tough even for sophisticated Web surfers to tell if their gear is being fingerprinted. Even if people modify their machines—adding or deleting fonts, or updating software—fingerprinters often can still recognize them. There's not yet a way for people to delete fingerprints that have been collected. In short, fingerprinting is largely invisible, tough to fend off and semi-permanent.

Device fingerprinting is legal. U.S. Rep. Bobby Rush (D.,Ill.), proposed legislation in July that would require companies that use persistent identifiers, such as device fingerprints, to let people opt out of being tracked online.

Fingerprinting companies are racing to meet the $23 billion U.S. online-ad industry's appetite for detailed consumer behavior. Previously, the companies focused on using device fingerprints to prevent software theft or to identify computers making fraudulent transactions, in hopes of preventing future attempts.

Mr. Norris's firm, BlueCava, this year spun off from anti-piracy company Uniloc USA Inc. to start offering services to advertisers and others. One of the leading e-commerce fraud-prevention firms, 41st Parameter Inc., has begun testing its device-fingerprinting techniques with several online-ad companies. Another anti-fraud company, iovation Inc. of Portland, Ore., says it is exploring the use of device profiles to help websites customize their content.

BlueCava says the information it collects about devices can't be traced back to individuals and that it will offer people a way to opt out of being tracked.

Still, Mr. Norris says it's tough to figure out how to alert people their devices are being fingerprinted. "We don't have all the answers, but we're just going to try to be really clear" about how the data is used, he says.

Neither BlueCava nor 41st Parameter explicitly notified the people whose devices have been fingerprinted so far. Both companies say the data-gathering is disclosed in the privacy policies of the companies they work with.

BlueCava says it doesn't collect personal information such as people's names. Its privacy policy says it gathers "just boring stuff that most people couldn't care less about."

Ori Eisen, founder of 41st Parameter, says using fingerprinting to track devices is "fair game" because websites automatically get the data anyway.

Some advertisers are enthusiastic about fingerprinting. Steel House Inc., a Los Angeles-based ad company, has been testing 41st Parameter's technology for three months on websites of its clients, which include Cooking.com Inc. and Toms Shoes Inc. (Clients weren't notified of the test, and fingerprints weren't used to display ads.)

In its examination of 70 million website visits, 41st Parameter found it could generate a fingerprint about 89% of the time. By comparison, Steel House was able to use cookies for tracking on only about 78% of visits, because some people blocked or deleted cookies.

"It's almost like a revolution," says Mark Douglas, founder and CEO of Steel House. "Our intent is that it can completely replace the use of cookies."

Steel House offers people a way to opt out of its current cookie-based ads and says it would do the same if it adopts fingerprints."I definitely don't want to be in the sights of the privacy people," Mr. Douglas says.

Computers need to broadcast details about their configuration in order to interact smoothly with websites and with other computers. For example, computers announce which specific Web browsers they use, along with their screen resolution, to help websites display correctly.

There are hundreds of parameters. "We call them the 'toys on the table,'" says Mr. Norris of BlueCava. "Everyone has the same toys on the table. It's how you rearrange them or look at them that is the secret sauce" used to fingerprint a specific computer.

BlueCava's secret sauce hails from Sydney, Australia, in the early 1990s. Back then, inventor Ric Richardson was helping musicians, including the band INXS, to use new software for playing their electronic keyboards.

"They'd say what sound they wanted, and I'd do it," says Mr. Richardson, who today works out of a van parked near an Australia beach.

Mr. Richardson was frustrated when he tried to sell the music software, because there was no way to let people test it before buying. So he designed a "demonstration" version of the software that would let people test it, but not copy it.

His idea: Configure his software to work only after it was linked to a unique computer. So, he developed a way to catalog each computer's individual properties. He found many subtle variations, among even outwardly similar machines.

"It was amazing how different they were," he says. "There are literally hundreds of things you can measure."

In 1992, he borrowed $40,000 from his parents, filed a patent application for a "system for software registration" and founded a company, Uniloc Corp.

This year, Uniloc started trying to broaden its business away from software-piracy prevention. It recruited Mr. Norris, then running a company that

provided photos for advertisers, to seek new uses for its technology. "What I saw was this different way of looking at things on the Web," Mr. Norris says.

Mr. Norris became CEO and spun off BlueCava to market device fingerprinting both to fraud-prevention and online-ad firms. Eventually, he hopes BlueCava can fingerprint everything from automobiles to the electrical grid. In October 2010, Texas billionaire Mark Cuban led a group of investors who put $5 million into BlueCava.

BlueCava embeds its technology in websites, downloadable games and cellphone apps. One of its first customers was Palo Alto, Calif.-based IMVU Inc., which operates an online game where 55 million registered players can build virtual identities and chat in 3-D. It wanted to combat fraudsters who were setting up multiple accounts to buy virtual clothing and trinkets with stolen credit-card numbers. Kevin Dasch, a vice president at IMVU, says BlueCava's technology "has led to a significant decline in our fraud rates."

Later this year, BlueCava plans to launch its reputation exchange, which will include all the fingerprints it has collected so far.

Unlike most other fraud-prevention companies, BlueCava plans to merge its fraud data with its advertising data. Rivals say they don't mix the two types of data.

Greg Pierson, chief executive of iovation, says the company will never disclose specific information about people's Web-browsing behavior, "because it's unnecessary and it's dangerous. It's close to spying."

Mr. Norris says collecting that data is "standard practice" in the online-ad business.

Mr. Dasch of IMVU says he doesn't mind fingerprints of IMVU customers being added to the exchange, provided that they don't contain personally identifiable information such as user names, and that his company can use other exchange data in return.

The idea behind BlueCava's exchange is to let advertisers build profiles of the people using the devices it has identified. For instance, BlueCava will know that an IMVU fingerprint is from someone who likes virtual-reality games.

Other advertisers could then add information about that user. BlueCava also plans to link the profiles of various devices—cellphones, for instance—that also appear to be used by the same person.

Blue Cava also is seeking to use a controversial technique of matching online data about people with catalogs of offline information about them, such as property records, motor-vehicle registrations, income estimates and other details. It works like this: An individual logs into a website using a name or e-mail address.

The website shares those details with an offline-data company, which uses the email address or name to look up its files about the person.

The data company then strips out the user's name and passes BlueCava information from offline databases. BlueCava then adds those personal details to its profile of that device.

As a result, BlueCava expects to have extremely detailed profiles of devices that could be more useful to marketers. In its privacy policy, BlueCava says it plans to hang onto device data "for the foreseeable future."

Advertisers are starting to test BlueCava's system. Mobext, the U.S. cellphone-advertising unit of the French firm Havas SA, is evaluating BlueCava's technology as a way to target users on mobile devices. "It's a better level of tracking," says Rob Griffin, senior vice president at Havas Digital.

Phuc Truong, managing director of Mobext, explains that tracking on cellphones is difficult because cookies don't always work on them. By comparison, he says, BlueCava's technology can work on all phones.

"I think cookies are a joke," Mr. Norris says. "The system is archaic and was invented by accident. We've outgrown it, and it's time for the next thing."

Published Dec. 1, 2010.

THE WALL STREET JOURNAL.
WSJ.com

How To Prevent Device Fingerprinting

BY JENNIFER VALENTINO-DEVRIES

It's extremely difficult for people surfing the Web to avoid being tracked by device-fingerprinting technology.

But those who would like to try to block fingerprinting can use two different, admittedly extreme, techniques: One method can slow Web browsing to a crawl, and the other forces many websites to display incorrectly.

When it comes to device fingerprinting, "we have no convenient options for privacy," said Peter Eckersley, staff scientist at the Electronic Frontier Foundation, a privacy-advocacy group. "All the things we can do are inconvenient to the point of being really impractical." In a study this year, Mr. Eckersley found that about 91% of nearly 1 million computer users surveyed could be fingerprinted simply by visiting a website.

Fingerprints are tough to avoid because they are lifted from data that are routinely passed from computers to websites automatically. This information includes things like the fonts installed on a machine and data about the computer's clock, down to the millisecond.

Even if a user changes one parameter, sophisticated fingerprinters can still recognize the machine. And unlike cookies, fingerprints don't leave a file on the user's machine, so there is nothing to delete.

However, there is one clue that a device is being fingerprinted. A small bit of software called JavaScript can be used to collect fingerprint data. JavaScript is ubiquitous, but if it asks a browser for certain types of information, that could mean a fingerprinter is at work.

Users intent on preventing device fingerprinting can block JavaScript. But that means some parts of a website—such as video and interactive graphics—may not load, resulting in a blank space on the page.

One of the most popular ways to block JavaScript is to use the Mozilla Foundation's Firefox browser with a small "add-on" program called NoScript, available at http://noscript.net/. This program stops JavaScript on pages and allows people to create "whitelists" so code on trusted pages is permitted.

However, fingerprinters can gather some data without using JavaScript. People worried about that situation can use a tool called Tor along with a related Firefox add-on called Torbutton. Tor, which often is used by dissidents in countries that censor Internet traffic, has technology that hides a user's Internet Protocol (IP) address. And Torbutton disables all plug-ins that can send information about your computer—including those that use JavaScript and other technologies. It also changes some of the information your browser discloses to make it appear less unique.

The engineers behind Tor and Torbutton also say they are working with browser makers to make it easier to develop protections against fingerprinting and other security issues. They hope their efforts will help curb the disclosure of data in HTML5, the upcoming version of the language used to code Web pages. In HTML5, some data can be disclosed independent of JavaScript.

The tools are available at https://www.torproject.org.

With such tools, though, users won't be able to watch Flash videos or do many things on websites, and their browsing is likely to be slower than normal.

And there are other caveats: The techniques don't block fingerprinters whose code is embedded in games that users download to their computers or apps on user's cellphones.

The companies known to be pursuing fingerprinting for advertising purposes are currently conducting live tests and collecting data; they say it would be possible to have an opt-out system for consumers at a later date, but they don't have such a system available yet.

Published Nov. 30, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Your Apps Are Watching You

BY SCOTT THURM and YUKARI IWATANI KANE

Few devices know more personal details about people than the smartphones in their pockets: phone numbers, current location, often the owner's real name—even a unique ID number that can never be changed or turned off.

These phones don't keep secrets. They are sharing this personal data widely and regularly, a Wall Street Journal investigation has found.

An examination of 101 popular smartphone "apps"—games and other software applications for iPhone and Android phones—showed that 56 transmitted the phone's unique device ID to other companies without users' awareness or consent. Forty-seven apps transmitted the phone's location in some way. Five sent age, gender and other personal details to outsiders.

The findings reveal the intrusive effort by online-tracking companies to gather personal data about people in order to flesh out detailed dossiers on them.

Among the apps tested, the iPhone apps transmitted more data than the apps on phones using Google Inc.'s Android operating system. Because of the test's size, it's not known if the pattern holds among the hundreds of thousands of apps available.

Apps sharing the most information included TextPlus 4, a popular iPhone app for text messaging. It sent the phone's unique ID number to eight ad companies and the phone's zip code, along with the user's age and gender, to two of them.

Both the Android and iPhone versions of Pandora, a popular music app, sent age, gender, location and phone identifiers to various ad networks. iPhone and Android versions of a game called Paper Toss—players try to throw paper wads into a trash can—each sent the phone's ID number to at least five ad companies. Grindr, an iPhone app for meeting gay men, sent gender, location and phone ID to three ad companies.

"In the world of mobile, there is no anonymity," says Michael Becker of the Mobile Marketing Association, an industry trade group. A cellphone is "always with us. It's always on."

iPhone maker Apple Inc. says it reviews each app before offering it to users. Both Apple and Google say they protect users by requiring apps to obtain permission before revealing certain kinds of information, such as location.

"We have created strong privacy protections for our customers, especially regarding location-based data," says Apple spokesman Tom Neumayr. "Privacy and trust are vitally important."

The Journal found that these rules can be skirted. One iPhone app, Pumpkin Maker (a pumpkin-carving game), transmits location to an ad network without asking permission. Apple declines to comment on whether the app violated its rules.

Smartphone users are all but powerless to limit the tracking. With few exceptions, app users can't "opt out" of phone tracking, as is possible, in limited form, on regular computers. On computers it is also possible to block or delete "cookies," which are tiny tracking files. These techniques generally don't work on cellphone apps.

The makers of TextPlus 4, Pandora and Grindr say the data they pass on to outside firms isn't linked to an individual's name. Personal details such as age and gender are volunteered by users, they say. The maker of Pumpkin Maker says he didn't know Apple required apps to seek user approval before transmitting location. The maker of Paper Toss didn't respond to requests for comment.

Many apps don't offer even a basic form of consumer protection: written privacy policies. Forty-five of the 101 apps didn't provide privacy policies on their websites or inside the apps at the time of testing. Neither Apple nor Google requires app privacy policies.

To expose the information being shared by smartphone apps, the Journal designed a system to intercept and record the data they transmit, then decoded the data stream. The research covered 50 iPhone apps and 50 on phones using Google's Android operating system. (See Methodology on p. xx.)

The Journal also tested its own iPhone app; it didn't send information to outsiders. The Journal doesn't have an Android phone app.

Among all apps tested, the most widely shared detail was the unique ID number assigned to every phone. It is effectively a "supercookie," says Vishal Gurbuxani, co-founder of Mobclix Inc., an exchange for mobile advertisers.

On iPhones, this number is the "UDID," or Unique Device Identifier. Android IDs go by other names. These IDs are set by phone makers, carriers or makers of the operating system, and typically can't be blocked or deleted.

"The great thing about mobile is you can't clear a UDID like you can a cookie," says Meghan O'Holleran of Traffic Marketplace, an Internet ad network that is expanding into mobile apps. "That's how we track everything."

Ms. O'Holleran says Traffic Marketplace, a unit of Epic Media Group, monitors smartphone users whenever it can. "We watch what apps you

download, how frequently you use them, how much time you spend on them, how deep into the app you go," she says. She says the data is aggregated and not linked to an individual.

The main companies setting ground rules for app data-gathering have big stakes in the ad business. The two most popular platforms for new U.S. smartphones are Apple's iPhone and Google's Android. Google and Apple also run the two biggest services, by revenue, for putting ads on mobile phones.

Apple and Google ad networks let advertisers target groups of users. Both companies say they don't track individuals based on the way they use apps.

Apple limits what can be installed on an iPhone by requiring iPhone apps to be offered exclusively through its App Store. Apple reviews those apps for function, offensiveness and other criteria.

Apple says iPhone apps "cannot transmit data about a user without obtaining the user's prior permission and providing the user with access to information about how and where the data will be used." Many apps tested by the Journal appeared to violate that rule, by sending a user's location to ad networks without informing users. Apple declines to discuss how it interprets or enforces the policy.

Phones running Google's Android operating system are made by companies including Motorola Inc. and Samsung Electronics Co. Google doesn't review the apps, which can be downloaded from many vendors. Google says app makers "bear the responsibility for how they handle user information."

Google requires Android apps to notify users, before they download the app, of the data sources the app intends to access. Possible sources include the phone's camera, memory, contact list, and more than 100 others. If users don't like what a particular app wants to access, they can choose not to install the app, Google says.

"Our focus is making sure that users have control over what apps they install, and notice of what information the app accesses," a Google spokesman says.

Neither Apple nor Google requires apps to ask permission to access some forms of the device ID, or to send it to outsiders. When smartphone users let an app see their location, apps generally don't disclose if they will pass the location to ad companies.

Lack of standard practices means different companies treat the same information differently. For example, Apple says that, internally, it treats the iPhone's UDID as "personally identifiable information." That's because, Apple says, it can be combined with other personal details about people—such as names or email addresses—that Apple has via the App Store or its iTunes music services. By contrast, Google and most app makers don't consider device IDs to be identifying information.

A growing industry is assembling this data into profiles of cellphone users. Mobclix, the ad exchange, matches more than 25 ad networks with some 15,000 apps seeking advertisers. The Palo Alto, Calif., company collects phone IDs, encodes them (to obscure the number), and assigns them to interest categories based on what apps people download and how much time they spend using an app, among other factors.

By tracking a phone's location, Mobclix also makes a "best guess" of where a person lives, says Mr. Gurbuxani, the Mobclix executive. Mobclix then matches that location with spending and demographic data from Nielsen Co.

In roughly a quarter-second, Mobclix can place a user in one of 150 "segments" it offers to advertisers, from "green enthusiasts" to "soccer moms." For example, "die hard gamers" are 15-to-25-year-old males with more than 20 apps on their phones who use an app for more than 20 minutes at a time.

Mobclix says its system is powerful, but that its categories are broad enough to not identify individuals. "It's about how you track people better," Mr. Gurbuxani says.

Some app makers have made changes in response to the findings. At least four app makers posted privacy policies after being contacted by the Journal, including Rovio Mobile Ltd., the Finnish company behind the popular game Angry Birds (in which birds battle egg-snatching pigs). A spokesman says Rovio had been working on the policy, and the Journal inquiry made it a good time to unveil it.

Free and paid versions of Angry Birds were tested on an iPhone. The apps sent the phone's UDID and location to the Chillingo unit of Electronic Arts Inc., which markets the games. Chillingo says it doesn't use the information for advertising and doesn't share it with outsiders.

Apps have been around for years, but burst into prominence when Apple opened its App Store in July 2008. Today, the App Store boasts more than 300,000 programs.

Other phone makers, including BlackBerry maker Research in Motion Ltd. and Nokia Corp., quickly built their own app stores. Google's Android Market, which opened later in 2008, has more than 100,000 apps. Market researcher Gartner Inc. estimates that world-wide app sales this year will total $6.7 billion.

Many developers offer apps for free, hoping to profit by selling ads inside the app. Noah Elkin of market researcher eMarketer says some people "are willing to tolerate advertising in apps to get something for free." Of the 101 apps tested, the paid apps generally sent less data to outsiders.

Ad sales on phones account for less than 5% of the $23 billion in annual Internet advertising. But spending on mobile ads is growing faster than the market overall.

Central to this growth: the ad networks whose business is connecting advertisers with apps. Many ad networks offer software "kits" that automatically insert ads into an app. The kits also track where users spend time inside the app.

Some developers feel pressure to release more data about people. Max Binshtok, creator of the DailyHoroscope Android app, says ad-network executives encouraged him to transmit users' locations.

Mr. Binshtok says he declined because of privacy concerns. But ads targeted by location bring in two to five times as much money as untargeted ads, Mr. Binshtok says. "We are losing a lot of revenue."

Other apps transmitted more data. The Android app for social-network site MySpace sent age and gender, along with a device ID, to Millennial Media, a big ad network.

In its software-kit instructions, Millennial Media lists 11 types of information about people that developers may transmit to "help Millennial provide more relevant ads." They include age, gender, income, ethnicity, sexual orientation and political views. In a re-test with a more complete profile, MySpace also sent a user's income, ethnicity and parental status.

A spokesman says MySpace discloses in its privacy policy that it will share details from user profiles to help advertisers provide "more relevant ads." My Space is a unit of News Corp., which publishes the Journal. Millennial did not respond to requests for comment on its software kit.

App makers transmitting data say it is anonymous to the outside firms that receive it. "There is no real-life ID here," says Joel Simkhai, CEO of Nearby Buddy Finder LLC, the maker of the Grindr app for gay men. "Because we are not tying [the information] to a name, I don't see an area of concern."

Scott Lahman, CEO of TextPlus 4 developer Gogii Inc., says his company "is dedicated to the privacy of our users. We do not share personally identifiable information or message content." A Pandora spokeswoman says, "We use listener data in accordance with our privacy policy," which discusses the app's data use, to deliver relevant advertising. When a user registers for the first time, the app asks for email address, gender, birth year and ZIP code.

Google was the biggest data recipient in the tests. Its AdMob, AdSense, Analytics and DoubleClick units collectively heard from 38 of the 101 apps. Google, whose ad units operate on both iPhones and Android phones, says it doesn't mix data received by these units.

Google's main mobile-ad network is AdMob, which it bought this year for $750 million. AdMob lets advertisers target phone users by location, type of device and "demographic data," including gender or age group.

A Google spokesman says AdMob targets ads based on what it knows about the types of people who use an app, phone location, and profile information a user has submitted to the app. "No profile of the user, their device,

where they've been or what apps they've downloaded, is created or stored," he says.

Apple operates its iAd network only on the iPhone. Eighteen of the 51 iPhone apps sent information to Apple.

Apple targets ads to phone users based largely on what it knows about them through its App Store and iTunes music service. The targeting criteria can include the types of songs, videos and apps a person downloads, according to an Apple ad presentation reviewed by the Journal. The presentation named 103 targeting categories, including: karaoke, Christian/gospel music, anime, business news, health apps, games and horror movies.

People familiar with iAd say Apple doesn't track what users do inside apps and offers advertisers broad categories of people, not specific individuals.

Apple has signaled that it has ideas for targeting people more closely. In a patent application filed in May 2010, Apple outlined a system for placing and pricing ads based on a person's "Web history or search history" and "the contents of a media library." For example, home-improvement advertisers might pay more to reach a person who downloaded do-it-yourself TV shows, the document says.

The patent application also lists another possible way to target people with ads: the contents of a friend's media library.

How would Apple learn who a cellphone user's friends are, and what kinds of media they prefer? The patent says Apple could tap "known connections on one or more social-networking websites" or "publicly available information or private databases describing purchasing decisions, brand preferences," and other data. In September 2010, Apple introduced a social-networking service within iTunes, called Ping, that lets users share music preferences with friends. Apple declined to comment.

Tech companies file patents on blue-sky concepts all the time, and it isn't clear whether Apple will follow through on these ideas. If it did, it would be an evolution for Chief Executive Steve Jobs, who has spoken out against intrusive tracking. At a tech conference in June 2010, he complained about apps "that want to take a lot of your personal data and suck it up."

Tom McGinty and Jennifer Valentino-DeVries contributed to this report.

Published Dec. 18, 2010.

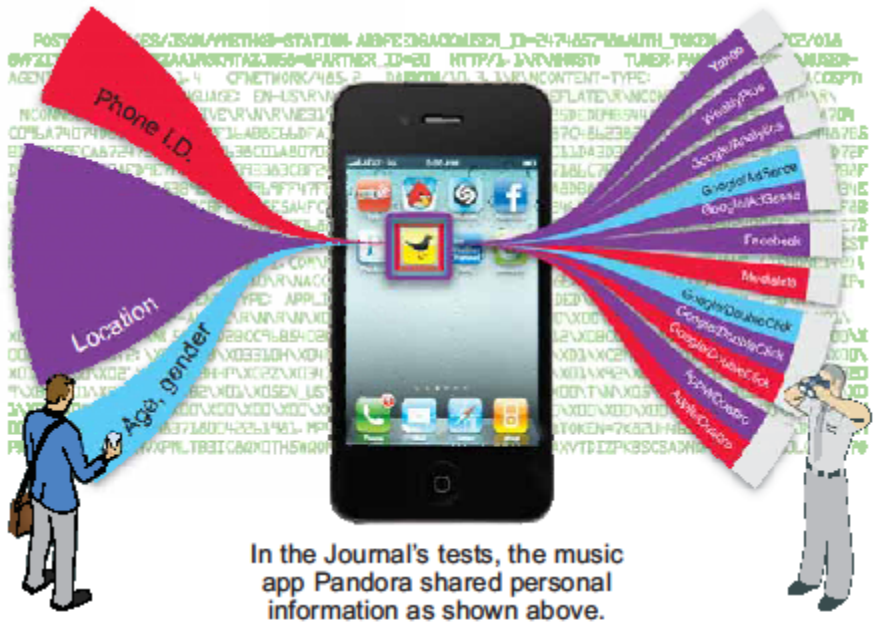
THE WALL STREET JOURNAL.
WSJ.com

Explore the Data

The personal data that your smartphone stores ...

... can be collected by the apps that you download ...

... and is often sent to outside companies.



In the Journal's tests, the music app Pandora shared personal information as shown above.

EXPLORE ALL THE APPS:

http://blogs.wsj.com/wtk-mobile/

THE WALL STREET JOURNAL.
WSJ.com

What Can You Do? Not Much

BY JENNIFER VALENTINO-DEVRIES

It's nearly impossible to prevent cellphone "apps"—games and other software—from transmitting information about a phone and its owner.

Turning off the phone's location services can restrict tracking by location. But it can limit some phone features like maps.

A few mobile marketing companies offer an "opt out" that prevents the use of tracking data to deliver targeted ads on websites viewed on cellphones. But most don't apply to apps. For instance, Ringleader Digital Inc.'s opt-out at ringleaderdigital.com/optout.php applies only to Internet browsing.

Ad company Jumptap Inc. says that opt.jumptap.com/optout/opt?jt, its opt-out, also doesn't apply to apps.

However, Jumptap says iPhone users can opt out of targeted ads in apps by emailing their Unique Device Identifier, or UDID, to optout@jumptap.com. The UDID is found by connecting the phone to iTunes and clicking on the serial number shown.

Apple Inc. says the opt-out for its mobile-ad system, iAd, does work for apps because it is tied to users' iTunes accounts rather than the Web browser. The opt-out (oo.apple.com) doesn't prevent iTunes data from being collected.

Google Inc. says it doesn't offer an opt-out for ads in apps because it doesn't create profiles of app users. It says its in-app ads aren't targeted based on user profiles.

Published Dec. 18, 2010.

THE WALL STREET JOURNAL.
WSJ.com

What Settings to Look For in Apps

BY JENNIFER VALENTINO-DEVRIES

Smartphone applications are passing along a trove of data to marketing companies, a study by The Wall Street Journal has found. There isn't much consumers can do about it — there's no way to block or detect much of the tracking technology used by apps on mobile devices.

The most important thing a user can do is pay attention to the information each app is requesting.

Apple Inc.'s iPhones generally request users' permission before apps can access the phone's location information. iPhone users can block the phone from sending location data by going to "settings" and "general."

People who have iOS 4, Apple's latest operating system for mobile devices, also can block an individual app's access to location by clicking on "location settings" and scrolling down the list of apps. Programs that transmitted location data in the past 24 hours will be marked with a little arrow on that list.

Apple's agreement for software developers says that apps "may not collect user or device data without prior user consent." There are no tools for users to determine whether an app is following that agreement.

Devices running Google Inc.'s Android operating system are required to tell users much of the information they gather before the person installs the software. Apps don't have to disclose whether they are collecting the Android ID, a number than uniquely identifies the phone.

If users don't want to grant the app access to that data, they shouldn't download the program, Google says. Users can't block apps they download from accessing that data.

Many Android apps request dozens of types of information, making the choice seem overwhelming. Here are some important items to focus on.

- In the list of permissions, the section marked "your personal information" is important and can include things like "read contact data" and "read Internet's history and bookmarks."
  Some apps, such as those that save Web pages to be read offline, might legitimately need this access. But if an app is requesting such data without a clear need, people might think twice about using it.
- Under the heading "phone calls," many apps want to "read phone state and identity." This is common because the ability to "read

phone state" is what lets the app know when the phone has an incoming call.

But the same permission allows the app to see some of the phone's unique numerical identifiers. So users should pay attention to this request and avoid untrusted applications that include it.

- Another important section: "your location," where there are two options. "Fine (GPS) location" uses Global Positioning System satellites to determine the phone's coordinates. "Coarse (network-based) location" means location is determined using things like Wi-Fi or cellular-tower signals.

  In the Journal's tests, "coarse" location frequently turned out to be more precise than "fine" location. The Toss-It Android application using "coarse" location derived latitude and longitude coordinates that Google Maps estimated to be 26 feet from the Denver office of the Journal's contractor.

  You can turn off location altogether by going to "menu," "settings," "location" and "location setting." Turning this feature off means important applications like maps won't work.
- Under the "system tools" heading, "read system log files" means the Android app could access files related to how the person uses applications on the phone. Programmers can use this to find software bugs. But in a note to developers, Google says the "entries can contain the user's private information."
- If the app has access to the items above and also asks for "full Internet access," that means it can send out the information via the Internet anytime it has a connection.

Published Dec. 19, 2010.

THE WALL STREET JOURNAL.
WSJ.com

Methodology

THE WALL STREET JOURNAL.
WSJ.com

Tracking the Trackers: Our Method

To determine the prevalence of Internet tracking technologies, The Wall Street Journal analyzed the 50 most-visited U.S. websites, as ranked by the comScore Media Metrix report from October 2009.

The Journal hired a technology consultant, Ashkan Soltani, to analyze the 50 sites for three types of tracking methods commonly used online: "HTML cookies," "Flash cookies" and "beacons."

HTML cookies are small text files, installed on a user's computer by a website, that assign the user's computer a unique identity and can track the user's movements on a site. Flash cookies are used in conjunction with Adobe Systems' Flash software, which is widely used to display graphics and video on websites. Beacons are bits of software code on a site that can transmit data about a user's browsing behavior.

Mr. Soltani visited the 50 sites between Dec. 10, 2009, and Jan. 14, 2010. Before each session, he cleared his computer of all browser data, including HTML cookies, Flash cookies and beacons. Each session consisted of visiting 20 pages per site. In one case, involving PayPal, he visited only six pages because viewing more would have required logging in to the PayPal service.

Mr. Soltani used Mozilla Firefox 3.5 and Adobe Flash Player 10.0. Following each session, he examined the tracking files that had been placed on the computer.

Beacons typically don't place a file on a computer. To trace them, Mr. Soltani used Ghostery, a small piece of software that can tell if a beacon is sending information from the website being examined.

Mr. Soltani also used a network-analyzer program to record all communication during a session, and to identify when his computer connected to other sites, to download an ad, for example.

At the time of his hiring by the Journal, Mr. Soltani was an independent consultant. For his master's thesis at the University of California, Berkeley, he and co-authors analyzed the use of beacons at the top 100 U.S. websites. He is now a contract technologist at the Federal Trade Commission. The FTC had no role in this study.

The Journal database also contains information collected by PrivacyChoice LLC about the privacy policies of companies that place these tracking files on websites. PrivacyChoice, founded by tech entrepreneur Jim Brock, provides privacy-consulting services to websites and doesn't accept money from ad companies that it surveys.

PrivacyChoice also provided the technology for the TrackerScan software that The Wall Street Journal is offering to readers to determine what cookies and other tracking tools are present on their own computers. You can access the software at WSJ.com/wtk.

The Journal compiled an "exposure index" for the 50 sites it examined, combining Mr. Soltani's findings with PrivacyChoice's analysis of cookie-placers, to determine how much each site exposes visitors to intrusive monitoring.

The exposure index gives each site a score based on eight criteria in PrivacyChoice's analysis: whether the site belongs to an industry self-regulating group; whether it lets users opt out of receiving cookies; whether it is part of an advertising or tracking network; whether it shares data it collects with others; whether it promises to keep user data anonymous; how long it retains user data; and how it handles sensitive data such as financial or health information.

A site's exposure index is the sum of the scores for each cookie, beacon and Flash cookie found on that site. The Journal used statistical analysis to group the 50 sites into four clusters of sites with generally similar characteristics.

THE WALL STREET JOURNAL.
WSJ.com

How the Analysis of Children's Websites Was Conducted

To determine the prevalence of Internet tracking technologies, The Wall Street Journal analyzed 50 of the most-visited U.S. websites for children and teens, as ranked by the comScore Media Metrix report from April 2010. The Journal excluded sites it had analyzed in its earlier database online.wsj.com/article/SB10001424052748703977004575393173432219064.html of major websites, and sites where fewer than 25% of visitors are under 18, according to comScore.

The Journal hired a technology consultant, David Campbell, to analyze the 50 sites for three types of tracking methods commonly used online: "HTML cookies," "Flash cookies" and "beacons." Mr. Campbell is a principal at Electric Alchemy in Denver, which specializes in software security and information assurance.

HTML cookies are small text files, installed on a user's computer by a website, that assign the user's computer a unique identity and can track the user's movements online. Flash cookies are used in conjunction with Adobe Systems' Flash software, which is widely used to display graphics and video on websites. Beacons are bits of software code on a site that can transmit data about a user's browsing behavior.

Mr. Campbell visited the 50 sites between June and August 2010. Before each session, Mr. Campbell cleared his computer of all browser data, including HTML cookies, Flash cookies and beacons. Each session consisted of visiting 20 pages per site.

Mr. Campbell used Mozilla Firefox 3.5 and Adobe Flash Player 10.0. Following each session, he examined the tracking files that had been placed on the computer.

Beacons typically don't place a file on a computer. To trace them, Mr. Campbell used Ghostery, a small piece of software that can tell if a beacon is sending information from the website being examined.

Mr. Campbell also used a network-analyzer program to record all communication during a session, and to identify when his computer connected to other sites, to download an ad, for example.

The Journal database also contains information collected by PrivacyChoice LLC about the privacy policies of companies that place these

tracking files on websites. PrivacyChoice, founded by tech entrepreneur Jim Brock, provides privacy-consulting services to websites and doesn't accept money from ad companies that it surveys.

PrivacyChoice also provided the technology for the TrackerScan software that The Wall Street Journal is offering to readers to determine what cookies and other tracking tools are present on their own computers. (You can access the software at WSJ.com/WTK.)

The Journal compiled an "exposure index" for the 50 sites it examined, combining Mr. Campbell's findings with PrivacyChoice's analysis of cookie-placers, to determine how much each site exposes visitors to intrusive monitoring.

The exposure index gives each site a score based on eight criteria in PrivacyChoice's analysis: whether the site belongs to an industry self-regulating group; whether it lets users opt out of receiving cookies; whether it is part of an advertising or tracking network; whether it shares data it collects with others; whether it promises to keep user data anonymous; how long it retains user data; and how it handles sensitive data such as financial or health information.

A site's exposure index is the sum of the scores for each cookie, beacon and Flash cookie found on that site. The Journal used statistical analysis to group the 50 sites into four clusters of sites with generally similar characteristics.

THE WALL STREET JOURNAL.
WSJ.com

The Journal's Cellphone Testing Methodology

The Wall Street Journal analyzed 50 popular applications, or "apps," on each of the iPhone and Android operating systems to see what information about the phones, their users and their locations the apps send to themselves and to outsiders.

The Journal selected the apps from the "most popular" lists on Apple's App Store and Google's Android Market as of mid-October 2010. (It also tested the newspaper's own iPhone app.)

The Journal hired a technology consultant, David Campbell, to analyze the apps. Mr. Campbell is a principal at Electric Alchemy in Denver, which specializes in software security.

Mr. Campbell used an iPhone 3G and a Samsung Captivate to test apps in an isolated network environment designed to capture all communications. The phones were restricted to run only a single application at a time to eliminate contamination and interference.

Mr. Campbell downloaded each app and used it for approximately five minutes to simulate normal use. For apps that asked, Mr. Campbell granted permission to transmit his location. For some apps, he created user profiles to see if details were transmitted to third parties. He asked some apps to search for friends. Some apps connected to the Internet, temporarily taking Mr. Campbell out of the app.

Mr. Campbell disabled the phones' cellular service and forced all the data traffic through a Wi-Fi connection, where it could be collected and analyzed. He used an open-source tool called "Mallory" to decrypt encrypted data.

THE WALL STREET JOURNAL.
WSJ.com

Glossary

Ad exchange—An auction-based marketplace where advertisers can bid to place ads in the space offered by websites.

Ad network—A company that sells ads on behalf of website publishers.

Aggregated information—Data combined from many individual users that can't identify anyone personally.

Anonymous information—Facts about you that don't identify you personally, such as age group and gender.

Beacons—Invisible software on many websites (also known as "bugs" or "pixels") that can track web surfers' location and activities online. Some are powerful enough to know what a user types on a particular site.

Behavioral targeting—Advertisers and websites use information about where you browse and what you search for online to guess your interests and decide what ads to show you. It's also called interest-based advertising or customized ads.

Cookie—Tiny text file put on your PC by websites or marketing firms that—depending on its purpose—might be used simply to remember your preferences for one site, or to track you across many sites.

Data exchange—A marketplace where advertisers bid for access to data about customers. Marketers then use this data to target ads. For example: A Denver hotel might bid to reach people known to have researched Denver hotels recently.

Exposure index—The Journal's analysis of how exposed your data is when you visit a website that has trackers. Each tracker was given a score based on how the tracking company collects, shares, and uses your data. A website's exposure index was calculated using the sum of the scores of all of the trackers we found on that site.

First-party tracking file—Typically a cookie installed on your computer by a website for benign purposes such as keeping you logged in to that one site.

Flash cookie—Small file put on your computer by Adobe's Flash software, which is used by many sites to display video or ads. Flash cookies can be designed to re-install regular cookies that were previously deleted.

Internet Protocol (IP) address—A unique number assigned to every computer connected to the Internet. Any website you visit can know your IP address and through that can often know your general location.

Offline data—Information about you that comes from sources other than the Internet. It could include your zip code, estimated household income, the cars you own or the purchases you've made in a store.

Personally identifiable information—Data identifying you uniquely, such as your name, Social Security number, address or credit-card information.

Privacy policy—A notice on a website that discloses some or all the ways the site collects or uses information.

Third-party tracking file—A cookie, beacon or other tracking technology installed on your computer by an ad network or research firm that can track your activities across many websites.

Tracking company—Companies that use cookies and other tracking technology to collect online data about you.

User profile—Information about your actions, interests and characteristics that tracking companies compile about you.

Published July 31, 2010.