Testimony of Edward W. Felten
Professor of Computer Science and Public Affairs, Princeton University

United States House of Representatives, Energy and Commerce Committee
Subcommittee on Communications, Technology and the Internet, and
Subcommittee on Commerce, Trade, and Consumer Protection
Hearing on
Behavioral Advertising: Industry Practices and Consumers' Expectations
June 18, 2009

Chairmen Boucher and Rush, Ranking Members Stearns and Radanovich, and members of the committees, I thank you for the opportunity to testify about the technology behind behavioral advertising.

My name is Edward W. Felten. I am a Professor of Computer Science and Public Affairs at Princeton University. I also serve as the founding Director of the Center for Information Technology Policy, an interdisciplinary research and teaching center at Princeton that focuses on public policy issues relating to computers and the Internet. My primary background is in computer science, and my main subfields of computer science include computer security and privacy, and Internet technologies. I have served as an advisor or consultant to the U.S. Departments of Defense, Homeland Security, and Justice, and the Federal Trade Commission. I have testified twice previously before House hearings and once before a Senate hearing. I am a Fellow of ACM, the leading professional society for computer scientists, and I serve as Vice-Chair of USACM, which is ACM's U.S. Public Policy Council.

I have been asked to testify about technical aspects of behavioral advertising, such as how online ads are delivered, and how information can be gathered and used by ad services. In discussing these topics, it is important to distinguish between what the technology allows, and what ad services actually do. Responsible ad services typically collect less information, and track users less intensively, than the technology would allow. I will describe what is possible technically, and I will leave it to other witnesses to describe which of these technical possibilities their services exploit.

1. Ads and Privacy

The most serious privacy concerns are raised not by the presence of advertising, but by the gathering of information about users that can be used either to target ads or for other purposes. The same kind of information-gathering can and does occur in contexts other than advertising. Regardless of when, where, and how information is gathered, there is no *technical* barrier to the resale of that information into secondary markets, or to the reuse of the information for other purposes. Accordingly, my technical discussion will be in the context of advertising, but it will be concerned mostly with how information about users can be gathered and used.

2. How Online Ads are Delivered

A typical online advertising scenario involves four parties. A *user* views a web page created by a *content provider*. The content provider leaves a space on the page for an ad, and an *ad service* chooses an ad to fill that space. An *advertiser* pays the ad service to place its ads, and the ad service pays the content provider for providing space.

Although the user sees the content and the ad together, they typically come from different places. The user's computer fetches the content, which has a blank space where the ad will be. The content comes with instructions that tell the user's computer where to go to get the ad, and where the ad should be placed on the page. The user's computer, following these instructions, connects to the ad service and asks it to provide an ad. The ad service decides which ad to provide; the chosen ad is delivered to the user's computer where it is displayed. Notably, the content provider and the ad service need not, and often do not, communicate directly during this process. Instead, the content provider simply causes the user's computer to fetch and display an ad from the ad service.

3. How Ad Services Gather Information

The ad service can increase the effectiveness of the ads it places, and thereby increase its revenue and the revenue of the content provider, by placing ads that are especially likely to be interesting and relevant to the user. Ads can be targeted based on context, on the user's past behavior, or on other information known about the user. For example, I might be shown ads for baseball tickets because I am reading a page about baseball, or because I recently purchased a baseball glove from a sports web site, or because the ad service knows that a sports magazine is regularly delivered to my home address. In general, the more information the ad service has about the context and the user, the more precisely it can target ads. Thus ad services have a natural incentive to collect and use detailed information about users.

Ad services can acquire information in three basic ways: they can get information from content providers, they can link users' activities across multiple sites, and they can use third-party commercial databases.

A. Getting Information From Content Providers

Ad services' first source of information is the content provider, which can supply information about the user, or about what the user is doing. For example, suppose I am viewing a newspaper story that is shown with a banner ad. The content provider (in this case, the newspaper) can tell the ad service anything it knows about me, including information that I provided when I signed up for an account with the content provider: it can reveal that I am male; or that I am a male in his forties living in New Jersey; or that I am a 46-year-old male, married with children and living in the 08540 area code; or that I am Edward W. Felten of Princeton, New Jersey, credit card number _____. Similarly, the content provider can supply the ad service with information about my interests: that I tend to read about national news, technology, and sports; that I read more about baseball than about hockey; that I often read Los Angeles Dodgers box scores; or (hypothetically) that I have recently shown interest in stories about cancer treatments.

In addition to information about the user generally, the content provider can supply information about what the user is doing at the moment. For example, a newspaper site might tell the ad network which story I am currently reading, or a travel site might tell the ad network that I am currently shopping for a train ticket to Washington.

The technical mechanism for passing this information from content provider to ad service is straightforward. Recall that when the content provider sends a web page to the user's computer, the page comes along with a command which the user's computer will carry out to request an ad from the ad server. The content provider can attach information to this command, and the ad service can retrieve that information when it receives the command. Additionally, information can be passed directly from the content provider to the ad service, through a back channel rather than going through the user's computer. In principle, any information known to the content provider can be provided in this way.

In some cases, the content provider may be the same organization as the ad service, or they may have a common corporate parent. For example, Facebook may serve ads for placement on content pages provided by Facebook itself; or Doubleclick, which is owned by Google, may serve ads on content pages provided by Google. In such cases, information might flow more easily from the content provider to the ad service, although the company may choose to impose internal controls on such flows, or even to maintain a strict "Chinese wall" between its content provider and ad service components.

B. **Linking Users' Activities Across Multiple Sites**

The second way an ad provider can gather information is by linking together actions by the same user across different web sites. In the course of a week I may visit many sites that show ads from the same ad service. If the ad service can determine that all of these visits came from the same person, then it can link together the information it gets from those visits, to create a more complete picture of who I am and what my interests are. For example, suppose I go to a social network site to discuss baseball, and that site uses a particular ad service. If that ad service knows that I am the same person who previously visited a weather site to look up Thursday's weather forecast for Washington, then the site can place on the social-network page an ad for tickets to a Washington Nationals game. This is possible because the ad service can link together the fact that I checked the Washington weather forecast on one site yesterday, with the fact that I discussed baseball on a different site today.

Of course, the information that is linked might be much more sensitive. For example, suppose (hypothetically) that last week I visited a news site and read several stories about cancer treatments. If the ad service can tell that the person who read this cancer information is the same person who checked the weather forecast and joined the baseball discussion, then it can add the (hypothetical) cancer information to its profile of me.

Further, if the ad service is able to link this information to another action that it knows was taken by Edward W. Felten of Princeton, New Jersey, such as a credit card transaction, then it will know that all of these actions were taken by me. The ad service could associate my cancer-related reading with my true name and identity, even if the website where I read the cancer stories did not know or reveal my identity.

There are several technical mechanisms that services can use to link together visits by a single user to different sites at different times. The most common such mechanism involves web "cookies." When the browser on a user's computer interacts with a service across the Internet, standard web technologies allow the service to provide a small piece of information, known as a cookie, that will be stored on the user's computer. Later, whenever the same browser re-connects to the same service, the browser will give the service a copy of the cookie. Services often give computers a cookie containing a unique number; if the same computer connects to the same service again, providing a copy of the cookie, the service can use the unique number to recognize that it has seen this computer before.

If an ad service can link together my various online activities, and if the ad service remembers all of the information about me that content providers have passed along, then the ad service can build up a profile of my online activities and interests. If any of the content providers passed along personally identifying information sufficient to convey my real-world identity, then the ad service will be able to connect its profile of my online activities to my real-world identity. There are no *technical* barriers to the ad service selling this information to third parties.

C. Using Third-Party Commercial Databases

The third way an ad service can gather information for targeting ads is by buying information from third-party providers such as consumer information databases. This is possible if the ad service knows the real-world identity connected to the online activities. If the ad service does know the identity, then third party services can provide a wealth of additional information, such as the user's demographics, family information, and credit history, which can be incorporated into the ad service's profile of the user, to improve ad targeting.

Of course, the fact that something is possible as a technical matter does not imply that reputable ad services actually do it. In practice, information gathering by ad services may be restrained by law, by self-regulation, by social norms, or by market pressures such as the desire of users to avoid sites that carry privacy risks. Services may choose not to gather, or not to use, certain kinds of information, or to gather and use only information that is less specific or less sensitive. I will leave it to other witnesses to describe the practices of their companies.

4. Web Beacons: Tracking Without Ads

Behavioral tracking can also happen on its own, without any advertisements. In this case, the same kind of cookie system is used to track user behavior and interests, just without displaying an ad. The data is still gathered, and it can still later be used to target future ads, or for any other purpose. When the tracking happens on its own, without any ad being displayed to the user, the tracking code is known as a "web beacon."

From the ad service's standpoint, a web beacon has the information-gathering power of an ad, but of course it lacks the ad content. The ad service can still use the information gathered by the beacon to build up its profiles about users, in order to improve ad targeting to those users later.

From the user's standpoint, though, the experience is very different: a web beacon, unlike an advertisement, leaves no outward mark on the web site the user sees. (Technically sophisticated users, or simple software, could find the beacon by examining the web page's code, but few users will do this.) As a result, users who wish to avoid being tracked will not know there is anything to opt out of, even if the ad network or other service that has placed the beacon does offer an opt-out mechanism.

5. Self-Help by Users

Users who are willing to engage in self-help, in an attempt to stop ad services from tracking them, have a limited ability to do so. Users have two main self-help strategies: they can try to block ads entirely; or they can try to stop ad services from linking together their actions across web sites.

A. Blocking Ads Entirely

The first self-help strategy tries to block ads entirely. This requires more than simply blocking visual presentation of ads. If the goal is to prevent ad services from gathering information about the user and his activities, then the user must prevent his computer from communicating with ad services. The typical approach is to establish a blacklist containing the network addresses of known ad services, and to prevent the user's browser from connecting to addresses that are on the blacklist. If the user's browser never connects to an ad service, then the ad service does not see what the user is doing and cannot link together the user's actions across different web sites.

If adopted widely, ad-blocking could potentially endanger the business models of content providers who rely on advertising revenue. Today, relatively few people use ad blockers. There are technical countermeasures that ad services could adopt, in an effort to defeat ad-blockers, but these countermeasures would be only partially successful.

It would be possible, in principle, to create an ad blocker that allowed ads from services that complied with certain specified privacy guidelines, while blocking ads from other ad services. This could give content providers a way to get some ad revenue, while protecting users against some privacy risks. Doing this would require relatively straightforward technical modifications to existing ad blocking tools.

B. Stopping Ad Services

The second self-help strategy tries to stop ad services from linking together the user's actions across different web sites. These approaches mainly revolve around controlling web cookies. Recall that ad services often use cookies to place a unique mark on a particular user's computer, so the ad service can recognize the same computer later (and can infer that that computer is likely under the control of the same user). By erasing an ad service's cookie, the user can try to stop the ad service from connecting new actions to the user's previous history. In addition, most current browsers provide some kind of "anonymous browsing mode" (or similarly named feature) in which the browser tries to avoid giving content providers any clues about the user's past browsing history. The theory is that anything the user does while in anonymous browsing mode will not be linkable to anything the user did before.

In practice, anonymous browsing modes are not airtight. With some technical effort, an ad service that chose to do so could still track a user over time, even if the user entered anonymous browsing mode, and even if the user manually deleted cookies from his computer. There are other ways, besides cookies, for a service to detect unique marks on a user's computer; examples include so-called "Flash cookies" (which are not really cookies) as well as methods that measure unique attributes of a browser or computer. The technical details are complicated, so I will not try to explain here the limitations of these self-help measures. Suffice it to say that, given the complexity of today's web technology, there are a great many ways for an ad service to leave a subtle mark on the user's computer that can be detected later, and that it is unlikely that users will be able to shut down all of these pathways to linkability. For practical purposes we should assume that ad services will be able to link together user's activities, if they exert enough technical effort.

Once again, the fact that ad services have the technical ability to do something does not mean that responsible web services actually do it. A user who has deleted cookies or entered anonymous browsing mode has made clear his wish not to have his activities linked over time. Whether an ad service takes action to override the user's wish is a separate question. I will leave it to other witnesses to describe what their services do.

6. Allowing Users to Opt Out

Some ad services allow users to opt out of their behavioral tracking. Opt-out can mean different things for different ad services. For one ad service, opt-out may mean that the service stops consulting user profiles when choosing ads, but still continues to add information to those user profiles, and to retain the profiles for other purposes. For another ad service, opt-out may mean that the service stops gathering new information about the user, but retains and keeps using the information it already had. For a third ad service, opt-out may mean that the service discards all of the information it has about the user, and does not gather more. Users cannot tell these cases apart, unless the service makes specific statements about what opt-out means.

Designing an opt-out mechanism can be a bit tricky. On the one hand, the user has asked not to be tracked. On the other hand, the ad service must have some way to recognize when an opted-out user is visiting. A standard approach is for the ad network to put onto the user's computer a generic "opt-out" cookie, which does not uniquely identify the user's computer but simply marks the computer as one whose user has opted out from that ad service. Because cookies are visible only to the service that created them, a separate opt-out cookie is needed for each ad service from which the user wants to opt out.

Another disadvantage of using cookies as opt-out markers is that if the user takes steps to delete the cookies on his computer, or to enter anonymous browsing mode (which makes the cookies temporarily invisible), the effect will be to hide the opt-out cookie, causing the ad service to think that the user has not opted out and therefore is willing to be tracked. In this instance, deleting cookies or entering anonymous browsing mode, steps that usually protect user privacy, will have the perverse effect of removing the opt-out cookie and thereby exposing the user to greater information-gathering. The result is that cookie-based opt-out mechanisms are not as "sticky" as we might expect, and the user might have to opt out

again. In addition, because cookies are attached to a browser on an individual computer, rather than to a person, a user who opts out on one computer may still be tracked if he uses a different computer later.

Creating a single site that offers "one-stop shopping" for users who want to opt out requires cooperation among many ad services. This is what the Network Advertising Initiative (NAI), represented at this hearing by Mr. Curran, is trying to do. A user who visits the NAI site can use a single NAI page to get opt-out cookies from nearly thirty ad services.

Technical steps are possible to make opt-out more comprehensive. One approach is to modify the user's web browser software so that ad services' opt-out cookies can be permanently fixed in place, regardless of whether the user deletes other cookies or enters anonymous browsing mode. Browser extensions can be created to do this for an individual ad service's opt-out cookie, as in Google's opt-out browser extension, or for many services' opt-out cookies, as in the TACO browser extension. Alternatively, large web sites which identify their users, such as social networks or email services, could help their users get and install the full spectrum of opt-out cookies. This could be made very easy for users: the user might check a single box which would cause the social network or email site to ensure that the full spectrum of opt-out cookies remains on the user's computer, whenever the user returns to the site. Many technical options are available to help users express their opt-out preferences.

In the end, opt-out mechanisms depend on the good behavior of ad services. Users can express their desire to opt out, but there is little if anything that users can do *technically* to force ad services to respect that desire. Users can resort to the self-help mechanisms described above, but these have limited efficacy. Ultimately users must rely on well-behaved ad services to keep their promises.

7. Conclusion

Citizens are rightly concerned about the possibility that commercial entities will build extensive profiles of who they are and what they do online. Ad services are not the only parties who can assemble such profiles, but large ad services do have a prime opportunity to build profiles, due to their relationships with many content providers who can pass along information about users, and due to the ad services' ability to connect the dots by linking together a user's activities across different web sites.

All of this is possible, as a *technical* matter, which is not to say that responsible ad services do all of it, or even most of it. Ad services may be restrained by law, by self-regulation, by social norms, or by market pressures. What is clear is that technology, by itself, cannot protect users from broad gathering and use of information about what they do online.

I am grateful to both committees for holding today's important hearing. Online behavioral tracking—whether it is undertaken for advertising or for other purposes—is an important aspect of life online, for businesses and consumers alike.

Behavioral Advertising: Information Flow