On Internet Privacy and Profiling
Senate Commerce Committee
June 13, 2000

Richard M, Smith
Internet Consultant
Brookline, Massachusetts
rms2000@bellatlantic.net

## Introduction

To begin with, I would like to first thank the Chairman and the Senate Committee on Commerce, Science, and Transportation for this opportunity to testify today on the issue of online profiling and its impact on consumer privacy. It is indeed an honor to be here.

My own background is that I have spent almost 30 years in the computer software business both as a software engineer as well as a business owner. I retired last September as the President of Phar Lap Software, Inc., a company I co-founded 14 years ago. Since leaving Phar Lap, I have worked as a consultant specializing in Internet security and privacy issues.

The issue of online profiling is very controversial. The reason is quite simple to understand. Most consumers are very bothered by the fact that companies are monitoring their Web surfing habits. In addition, consumers are almost never informed about these monitoring activities and have never been asked if it is okay. To many people who learn about online profiling for the first time, their first impression is that it is something right out of Orwell's *1984*.

In my testimony today, I will be focusing on two major areas. To begin with, I will talk about how data is collected by Internet ad companies for use in online profiles. To date, I do not think that ad companies have been totally straight with consumers with their data collection practices. The second area I want to talk about today is the lack of proper notice to consumers about online profiling. I will be using real-life examples of some of things that I have seen in my own use of the Internet.

Along the way, I want to also suggest an alternative to online profiling which is content-based targeting for banner ads. Content-based targeting is typically employed in the off-line world (newspapers, TV, and magazines). It is much more privacy friendly than online profiling because it requires no tracking of individual users as they surf the Internet. The most banner ads shown today are already using content-based targeting because it is easy to understand and favored by advertisers.

## How data is collected for online profiles

To begin the discussion of data collection practices of Internet ad companies, the best place to start looking is at Internet search engine sites. Everyone seems to have their own favorite search engine and mine happens to be AltaVista. It also turns out that the AltaVista site has business relationships with DoubleClick and Engage who both are also testifying here today.

Most people probably have noticed at one time or another that the banner ads that they see on a search results page are related to what they are searching for. This is no accident. AltaVista employs DoubleClick to show banner ads at the site. One of the services that DoubleClick provides for advertisers is the ability to "purchase" keywords at the site. When a company owns a particular keyword or phrase, their banner ads will appear of the search results page for the keyword or phrase. Keywords are typically purchased on a month-by-month basis. They can be purchased either on an exclusive basis or can be shared with other companies.

Exhibit A illustrates how some common keywords such as "sports cars" and "vacation homes" will show relevant banner ads at AltaVista. A version of Exhibit A is also available at my Web site that shows in real-time what banner ads are being shown for common keywords. This demonstration is available at:

http://www.tiac.net/users/smiths/commerce/avads.htm

Advertisers like keyword targeted ads because it is more likely that people seeing their ads will be interested in their products. DoubleClick and AltaVista also like keyword targeted ads because they can charge a premium for them. This premium is typically 2 to 3 times more than standard ads at AltaVista.

But what about the consumer? How do they feel about keyword-targeted ads? The answers are a bit more difficult to come by. When many consumers notice keyword-targeted ads for the first time they get a bit uncomfortable. They realize that someone is watching them as they search the Internet with AltaVista. Most folks do not like to be watched and one of the first association that comes to mind is *1984*. On the other hand, I think most people will agree that if they are going to see banner ads at Web sites, they might as well be relevant to their interests.

AltaVista did not help matters much, because until January of this year, they did not disclose to users that banner ads can be targeted to search phrases. They also have made mixed efforts in informing users about their relationship with DoubleClick. However, a savvy Web user today who reads the AltaVista privacy policy will learn both about keyword-targeted ads and DoubleClick.

So do keyword-targeted ads present a privacy problem for users? I personally do not think so. In the Yellows Pages, we see ads for car dealerships in the automobile section. The same is true with the search results page for "cars" at AltaVista. I believe that this type of content-based targeting is valuable to both advertisers and consumers. It is an example of good Internet marketing.

However, there still are the concerns of consumers that they are being watched when they see keyword-targeted ads. How can these concerns be addressed? The first part of the solution is to provide adequate notice to consumers about the practice. For example, some of the search engine companies are now disclosing this practice in their privacy policies. The real answer for consumers is to make it clear that that their search strings are *never* saved in a database. Except for keeping aggregate statistics on the popularity of keywords, people's search strings should be discarded. More about this issue shortly.

But how does DoubleClick know what ad to display for a search keyword in the first place? Very simply, AltaVista gives DoubleClick, everyone's search strings. The hand-off is done right on the search results page. A banner ad is displayed as a image, and the URL of image is specially constructed by AltaVista to include the search string. Here is what one of these banner ad image tags looks like for the search string "sports cars":

```
<IMG SRC="http://ad.doubleclick.net/ad/altavista.digital.com
/result_front;kw=sports+cars;cat=totext;ord=1804224227?"
border=0 height=60 width=468>
```

You will notice that the search string is embedded as the "kw" parameter in the image URL.

So DoubleClick is being sent everyone's search strings at AltaVista. Pretty obviously you can learn a lot about a person by observing what they are searching for on the Internet. The ad network companies have realized this also and invented the idea of online profiling. The basic concept is for the ad server computers of the ad companies to track over time what an individual is searching for and to provide relevant ads to according to their search history. These personalized banner ads can be shown whenever someone searches for a keyword that has not been purchased by an advertiser. These same personalized ads can also be shown at other Web sites in the same ad network.

However it is pretty cumbersome for an ad network to remember every little search string that someone has used. Such a list does not lend itself to quickly selecting an ad for a user. In general, an ad server must decide on what ad a user sees in about 1/100 of a second. So in order to meet this time constraint, Internet ad companies instead build profiles of people. A profile is a table that rates a person on their level of interest in particular subjects. A profile might contain up to a thousand different subjects areas. These subjects areas might include things like sports (golf, tennis, football, etc.), travel (US, Canada, Europe, etc.) and food (cooking, gardening, etc.). A person is then scored for each of these subject areas. A score is a percentage. Zero percentage meaning no interesting, while one hundred percentage means extremely interested. These scores are updated in real-time from search strings and other data.

Advertisers can then target groups of users by instructing an Internet ad network to show their ads to people who have certain characteristics in their profiles. For example, a ski resort may want to have their ads to be shown only to people who appear by their profiles to have a strong interest in skiing. The targeting might also be indirect. A car company might target ads for their luxury models at people who show an interest in European travel, while their middle-of-the-road models might be pitched to people who show an interest in American travel.

An online profile is created for a user the first time they are shown a banner ad from a particular Internet ad network. All of the scores in the profile are set to zero. The profile is stored at the ad server computers. It is updated in real-time according to the following information that is received by Internet ad networks:

- What search strings an individual searches for
- What Web pages an individual visits
- What banner ads an individual clicks on

A user can be tracked by an Internet ad company on any Web page that a banner ad appears that is served by the company.

In addition to their profile, a user is also assigned a unique customer ID number. This ID number is stored with the profile to identify who the profile belongs to. The ID number is also sent back to the user's computer as a cookie and stored on the hard drive of the computer. Then as the user surfs the Web and is shown more banner ads, this customer ID number is sent back to the Internet ad network with each and every request for a banner ad. The cookie is the mechanism that allows Internet ad networks to track people over time.

Cookies are anonymous in the sense that they do not say who a person is. However, personal information can be associated with a cookie and stored with a profile if a user provides this information to an Internet ad company. This is typically done using some sort of online contest or sweepstake where users are required to provide their names, addresses, and phone numbers. As an example, DoubleClick operates a Web site called NetDeals (http://www.netdeals.com) for this purpose.

In addition, using a technique called "cookie synchronization", it is possible for one Web site to provide an Internet ad network with personal and demographic data about users. Again this information can be associated with a cookie and stored in an online profile. Excite@Home is apparently using this technique to provide registration data to its sister company, MatchLogic, an Internet ad company.

On paper, the economic benefits of online profiling seem self-evident. In theory, a profiled banner ad should have an increased response rate because it is being better targeted. Advertisers can purchase a smaller number of ad impressions in order to get the same results. Ad networks can charge more money per ad impression because the higher perceived value.  Consumers are suppose to benefit because they will see less ads about products that they no interest in.

However in practice, the value of online profiling is yet to be proven.  The industry has not released any studies that show response rates are significantly higher for profiled ads.  In addition, the response rates need to go up more than the costs of profiling.  These costs include the premium paid for ads themselves plus the time it takes to figure out what profile works best for a particular ad.  This second point is very important.  It is unclear if advertisers can use all of the data that Internet ad companies can provide them.  This point was made recently in a *New York Times* article by Saul Hansell:

> ### So Far, Big Brother Isn´t Big Business
>
> http://www.nytimes.com/library/financial/personal/050700personal-privacy.html
> May 7, 2000
>
> "The few advertisers that have tried these systems have not yet given up on them. But most say the response to their ads does not go up enough to be worth the extra cost and bother. It seems easier for them to buy cheap shotguns, in effect, than expensive laser-guided rifles."

Regardless if online profiling systems make economic sense or not, from a privacy standpoint, they present some real dangers.  These systems are monitoring people as they surf Internet.  What data is being collected and what is being saved away is not made very clear.  All of the uses of this data is not disclosed and may change over time.  Also in spite of claims by Internet ad companies that the profiles are anonymous almost all of these companies maintain separate databases with personal data that can be combine with the anonymous profiles at anytime using cookie synchronization.

However the real danger that I see with online profiling is that Internet ad companies have set up extensive monitoring systems to provide data for profiling.  It is almost like they have put hidden microphones in our homes and our offices and they listening to what we do all day long.  Pretty obviously if you deploy hidden microphones, you are going to pick up information which is personal in nature.  And this is exactly what I have found on my own computer.  The data collection systems that the Internet ad companies are currently running are getting personal and sensitive information that almost everyone will agree is none of the business of these companies.  The problem here is one of collateral damage

**Data Spills**

The first problem that I have seen at many Web sites is the problem of data spills.  A data spill is where information that is typed into a form at a Web site is accidentally sent off to an Internet ad company.  Data spills are caused by poor Web site design  Because I do logging of my Internet traffic from my computer, I can detect data spills.  In a two-month period, I found close to 10 data spills of personal data to DoubleClick.  These data spills include things like my name, home address, Email address, and birth date.  Web sites that were sending off this data to DoubleClick included well-known sites like AltaVista, Real Networks, HealthCentral, Quicken, and Travelocity.

My Web site includes a write-up that describes how data spills occur in the first place and how they can be prevented. The URL of the write-up is available at:

http://www.tiac.net/users/smiths/privacy/banads.htm

In the write-up, I talk mostly about DoubleClick. They are going to be receiving the most information from data spills given that they are largest provider of banner ads. However, the problem can occur with any banner ad network and all companies are receiving this kind of personal data from Internet users. A recent example of data spill really illustrates the point. I found that on my computer the sign-up page for the contest Web site, Jackpot.com, gave away my Email address to three different companies all at the same time. The companies receiving my Email address were Flycast, YesMail, and Sabela. The Jackpot.com privacy policy states they never share personal data, but they seem to have a tough time keeping this promise. My enquiry to the company about the issue was answered with a denial that there was any problem. The customer support person simply repeated the claims of the privacy policy.

In general, Jackpot.com is the exception rather than the rule. Other Web sites have been more response and fixed the problems right away when I have brought them to their attention. In addition, in some discussions I have had with the Internet ad companies, they have made it clear that they do not want this of type of unsolicited personal information from users. However, from their perspective it is a problem they cannot directly solve because the issues are with the Web sites running the banner ads and not at the ad servers.

In the near term, I am hoping to see Internet ad companies publicly commit to not use this unsolicited personal data from data spills. The best place to do this I think is in their privacy policies. The idea here is to acknowledge the problem that Web sites may accidentally give away personal data, but the Internet ad networks will discard it and not make use it.

Over the long term, there is a simple technology solution to the problem that can be implemented by Web browser companies. This solution involves eliminating referring URLs for being sent in situations where a data spill is likely to occur. Referring URLs can contain the personal data in a data spill.

## Web Bugs

Besides banner ads, Internet Ad companies also track users with something I've nicknamed "Web Bugs". A Web Bug is an invisible image on a Web page that sends back the cookie of an Internet ad company to their servers. The main purpose of a Web Bug is to track what pages users are going to the Internet. Given that images are invisible on the page, the averagel user has no way of knowing that they are being tracked in this manner. In addition, to my knowledge, no Web site or Internet ad company has every disclosed the use of Web Bugs in their privacy policies.

Pretty obviously, people in the Internet ad business do not call these invisible images "Web Bugs'. Instead they use names like "clear GIFs", "1-by-1 pixels", "tracker GIFs", and sensors. Since no one has come up with a consistent name for them, I will continue to use the term "Web Bugs".

Even though there has not been very much public discussion about Web Bugs, they seemed to be employed by most Internet marketing companies. In my discussions with these companies, I have been told that they are used for these purposes:

- The see who has come to a Web site after viewing a banner ad
- To transfer both personal and non-personal information from a Web site to an Internet ad company
- To provide data to an online profile
- To count ad impressions and page hits

More technical information on Web Bugs can be found at my Web site at this URL:

http://www.tiac.net/users/smiths/privacy/wbfaq.htm

In addition, I have set up search page that will locate Web pages that employ Web Bugs. The page operates by giving special search string to AltaVista that has located the hidden images. The URL of the search page is:

http://www.tiac.net/users/smiths/privacy/wbfind.htm

The page will locate Web Bugs that have been placed around the Internet from more than 20 different Internet marketing companies

Although Internet ad companies represent that they do not do profiling of sensitive areas such as children, medical, financial, and sexual issues, most of them will use Web Bugs on pages that deal with these areas. Here are a few illustrations of Web pages that employ Web bugs that I believe most people will find troubling:

- Kids Zone of Santa.com (http://www.santa.com/santa/kidszone/index.htm)
- Procrit.com (http://www.procrit.com)
- Rodale Press (http://www.sexamansguide.com/a/home/order.rhtml)
- Metropolitan Life (http://metlife.com/Salescareers/Apply/Docs/online_interview.html)

The Procrit Web site is the most interesting use of Web Bugs on the list. Procrit is product of Ortho Biotech which is a subsidiary of Johnson and Johnson. The drug is used to fight anemia in patients with a number of different conditions including AIDS, cancer, and kidney disease. Hidden image files from DoubleClick are strategically placed on the Procrit Web site in order to distinguish if someone is at the site because they are interested in treatments because of AIDS vs. cancer vs. kidney disease. Needless to say, I believe that most visitors to the Procrit site would be very surprised to learn they are being monitored in this way. However, unless someone understands HTML source code and knows where to look, they would never see the Web Bugs at the site.

Web Bugs appeared to be employed by all of the Internet ad companies. AltaVista has found more 30,000 placed by DoubleClick and about 1,000 placed by Engage. Be Free, another Internet marketing company, has more a half of a million according to AltaVista.

Personally I am surprised that Web Bugs are ever used. When discovered, they undermine people’s trust in Web sites. Some sites I know have stopped using Web Bugs when they received enquires from the press and consumers about their presences on the sites. Two such sites were Nabisco Kids and the United States Air Force. Web Bugs are also playing a role in a number of the privacy lawsuits that have been filed against Web site and Internet ad companies.

The problem that I see with Web Bugs is that supply information on the sly to Internet ad companies that can be used in personal profiles. Given that this tracking is being done with no notice or consent, I find use of Web Bugs very problematic.

**Notice and Banner Ad Networks**

I want to shift gears for a second and talk about the problem of notice with online profiling. Most consumers are unlikely to be aware that they are being tracked as they surf the Web. I suspect that most consumers would be surprised that their computers are sending back information to Internet ad companies about what articles and Web pages they are reading online. They would probably also be more even dismayed to learn that some of this information actually is being used for profiling purposes. Most consumers are in the frame of mind that Web is just like other media such as television or newspapers. Reading an article in a newspaper is obviously anonymous unless a person chooses to tell someone else about what they have read. However, reading the same article in the online world can be very different. Two or three different companies may know what article someone has read, how long the article took to read it, and where the person went on the Web when they were done.

Over the last 3 or 4 years, the industry has settle on the use of Web site privacy policies to inform consumers about what data is being collected by a Web site and what is done with the data. Today almost all popular Internet sites have privacy policies in places. In most areas these privacy policies do an acceptable job of inform a consumer what they can expect with information. One very notable exception is the use of online profiling at their sites.

In addition, all of the major Internet ad companies also have privacy policies that describe how banner ad networks work, what data is being collected by these networks, and the details of online profiling. Also, most of the Internet ad companies offer an "OPT-OUT" to allow consumers the ability to turn off tracking and profiling.

However, there is one major flaw with the privacy policies of Internet ad companies. Consumers have almost no way of ever seeing these privacy policies. The problem here is the Internet ad companies are hidden in the background at Web sites and consumers by and large do not know anything about the companies. Web sites, in the own privacy policies, have not helped the situation very much for consumer. Although a Web site privacy policy may talk some about the Internet ad company they use, Web sites almost never link to the privacy policy of ad networks. For example, the AltaVista search engine finds less than 150 links to DoubleClick's privacy policy. Yet, DoubleClick has more than 12,000 Web sites that they provide banner ads for. A similar situation exists for Engage, less than 100 links are found to the Engage privacy policy, yet Engage and its sister companies provide banner ads for more than 6,000 sites.

There clearly is a problem here of Internet ad companies providing proper notice about online profiling.

## Conclusion

The bottom line for me on online profiling is that Internet ad companies are getting too much data about us. Their ad networks function as tracking systems the gather data about us from search strings, banners ads on Web pages we visit, data spills, and Web Bugs. Clearly the data collection systems of the Internet ad companies are gathering more information about us than is necessary to show banner ads.

I know that many people involved in regulation issues around Internet advertising support the concept of OPT-OUT from online profiling. At the present time, I feel extremely uncomfortable with OPT-OUT for the following reasons:

- It is nearly impossible for consumers to learn about how they can OPT-OUT to online profiling because of lack of almost any kind of reasonable notice about online profiling.
- Invisible Web Bugs can provide data to the online profiles and consumers have no method of knowing that they are being tracked.
- Data spills are providing personal data about users to Internet ad companies and the industry has taken no public steps to stop the problem
- Many of Internet ad companies have divisions or sister companies that maintain databases of personally identified data that can be combined with the anonymous profiles at any time.

I want to conclude my testimony with one quick statistic from my own travels around the Internet. As I mentioned earlier, I run software on computer that logs all of my transactions on the Internet. The last 6 months, I had about 250,000 Web transactions total. More than 10% of these transactions were with DoubleClick. This works out to about 150 transactions per day. This means that DoubeClick is receiving 150 URLs of

Web pages I am visiting each and everyday. In the offline world, I cannot think of one company that it is getting this amount of data about me. Not my phone company, not my bank, and not my credit card company.

Thank you again for this opportunity to address the Senate Commerce Committee.