Term Paper

Personalization vs. Privacy

Questions on Ethics in Software Engineering Raised by the DoubleClick / Abacus Direct Merger

Matthias Book

Table of Contents

Abstract

This paper describes the process of delivering targeted advertising messages to World Wide Web users and the resulting privacy concerns. After introducing the underlying cookie technology and advertising concepts, the paper discusses the implications of the merger of online advertising agency DoubleClick, Inc. and direct mail services provider Abacus Direct Corp. which gives DoubleClick the potential to associate anonymous user profiles with actual names and addresses. The paper finally presents the self-regulatory principles for online preference marketing established by the Network Advertising Initiative and closes with notes on the effectiveness of banner advertising.

Introduction: Personalization and Cookies

The most powerful technique in web publishing is also the most subtle: Instead of serving the same static page to their whole audience, many web sites today customize content for individual users, putting pages together dynamically as users request them. Often, users do not even realize they are seeing a page tailored specifically to them because the dynamic content is integrated seamlessly with static elements. This technique, known as *personalization*, is what sets the World Wide Web apart from any other broadcast medium.

The applications for personalization are virtually limitless: A news site, for example, can display headlines of unread articles prominently on the page, while removing articles which the user already read to reduce clutter. Taking this one step further, the site can record which articles the user reads and which he ignores in order to display a more relevant selection of headlines when he visits the next time. The same mechanism can be applied to products offered on an e-commerce site, advertising displayed on a content site, and many more.

The technical challenge of personalization lies in the fact that in order to serve customized content to a user, the system has to identify him first. This can be achieved by prompting the user to enter a login name before accessing any content. However, that solution is cumbersome for both the user and the system developer. The user will soon become irritated for having to enter a code every time he wants to use the system – while this is acceptable when security is a concern, it is not reasonable for looking up today's headlines. On the server side, the developer has to make sure that the system recognizes the user throughout his session although he enters his identification only once. Since HTTP is a stateless protocol [FGM99], successive requests cannot be associated with a specific session. Thus, the developer must build state management into the system at a higher level, for example by attaching the user's ID to every URL requested – a complicated and error-prone process.

In 1996, Netscape Communications introduced a feature in the Navigator browser that made the simulation of state in HTTP transactions much easier: A so-called *cookie* [NC96] is a piece of server-generated state information which is stored on the client and automatically transmitted back to the server with every client request. This mechanism has two advantages: Firstly, the developer does not need to implement any state management but can rely on the client to supply the necessary information. Secondly and more importantly, a cookie can be configured to be persistent on

the client even if the session with the server ends (i.e. the user leaves the web site or closes the browser). Yet, the next time the client connects to the server, it transmits the cookie again, thereby re-establishing the state that existed before. This way, if the cookie contains a unique identification value, the server can recognize the user without having to prompt for manual identification. The whole process is transparent to the user.

While many users are concerned with the privacy and security implications of letting web servers store and read data on their hard drives, the cookie mechanism implements some safeguards against abuse: Firstly and most importantly, the browser only transmits a cookie if the URL of the requested file matches the valid domain and valid path parameters specified when the cookie was generated. For example, if the valid domain was set to `.matthiasbook.de` and the valid path was set to `/papers` when generating a certain cookie, browsers would only send this cookie when they requested files in the `papers` path of the domain `matthiasbook.de` (including requests for files like `polaris.matthiasbook.de/papers/personalization/index.html`). If no valid domain or path are specified when a cookie is generated, the browser uses the host name of the server and the path to the file that generated the cookie. This mechanism ensures that a cookie can only be read by the server it was intended for, but not by any other server. As a second safeguard, transmission of a cookie can be restricted to secure channels, i.e. it will only be sent if an encrypted transmission protocol like HTTP over Secure Socket Layer (HTTPS) is used. This option is useful if the cookie contains confidential information like a password.

Online Preference Marketing and Network Advertising

The web advertising industry soon recognized cookies as an important tool for targeting users, i.e. for delivering the right ad to the right user. Before cookies, advertisers could reach their target audience only by placing ads on content sites that revolved around a topic related to their product. For example, a computer hardware manufacturer might place ads in the online edition of a computer magazine. This strategy has been successfully used in offline media for decades, and advertisers are always looking for ways to optimize ad placement – like placing an ad for a new printer right next to an article discussing printers.

Through the cookie mechanism, the web opens up a whole new dimension of fine-tuning parameters: By tracking which articles a user reads and which he ignores on a magazine site, the system can compile a profile of his interests. Based on that profile, the system then can deliver not only relevant headlines, but also relevant ads to the user.

Instead of targeting broad audiences, advertisers can now target individual users based on their preferences, a strategy known as *online preference marketing* (OPM).

At first sight, the cookie mechanism's safeguards seem to restrict the reach of OPM to single web sites: Since cookies cannot be read by different servers, a user profile compiled on one site cannot be associated with that user when he visits another site. However, that restriction can be circumvented when both sites use the same server to deliver ads, and the ad server generates and reads all the cookies. The key here is that cookies are not tied to whole web pages, but to individual files; and that web pages residing on one server can contain files from other servers. So, if pages on server A and server B both contain image files (like banner ads) taken from server C, then server C can read and write cookies associated with those image files every time a user views a page on server A or B.

Taking this idea to the real world, A and B would be content publishing sites like online magazines or search engines, and C would be an advertising agency that poses as an intermediary between advertisers and publishers. Since the clients always exchange cookies with the agency's server, regardless of which publishing server they are visiting, the agency can track users across multiple unrelated sites and compile preference profiles with data from all those sites. For example, if a user occasionally clicks on banner ads for scanners when visiting a hardware site, the agency might show him ads touting image processing applications the next time he visits a software site.

The more publishing sites work together with the agency, the greater the synergy effects of this *network advertising* model: User profiles become more and more detailed, and they can be used by more and more sites at the same time. This is a win-win-win solution for advertisers (because they can target users more precisely), publishers (because they can rely on an existing system with more detailed data than they could provide themselves), agencies (because everybody wants to get on their system) – but what about the users?

Personally and Non-Personally Identifiable Information

Many users feel their privacy is being violated by the information compiled in the profiles. Indeed, online advertising agencies use a broad spectrum of parameters to target ads to individual users. DoubleClick, Inc., the largest agency serving a network of 11,500 sites, uses the following information from the client to serve an ad [DC00a]:

- IP address
- top-level domain
- browser type and version
- operating system
- Internet service provider
- local time and date
- language
- page viewed

This information is transmitted by the browser every time it requests a file from any server since it is required by the HTTP specification. In addition, DoubleClick might also use demographic information provided by publishers or advertisers, such as gender, age, education, financial status, etc. With this data, DoubleClick enables advertisers to target users by a combination of various criteria [DC00b]:

- interest category
- search keyword
- editorial keyword
- geographic location
- top-level domain
- page viewed
- organization type, name, size and revenue
- Internet service provider
- operating system
- browser type
- day and time
- publisher-supplied demographic information

While these criteria allow for relatively fine targeting of ads, they all rely on *non-personally identifiable information* (Non-PII): Although DoubleClick associates individual users with specific profiles, the users still remain anonymous to the advertiser, agency and publisher since they don't know their name and address.

However, in 1999, DoubleClick acquired Abacus Direct Corp., a direct-marketing services company that maintains a database of names, addresses, retail purchasing habits and demographic information on 90% of American households. This move enables the company to associate the so-far anonymous profiles with the actual names and addresses of users. The process of correlating DoubleClick's existing non-PII with the *personally identifiable information* (PII) in the Abacus database is relatively easy: When a user visits a site in the DoubleClick network that requests personal identification (like name and address for online shopping), that site can send the personal data to DoubleClick where it is associated with the user's invididual cookie. DoubleClick can also look up the user's name in the Abacus database to find information on his offline buying behaviour. The combined PII and non-PII of each identified user is then stored in the Abacus Online database [DC00c].

By matching users' web activities with the names in the Abacus database, targeting on an unprecedented scale is possible: According to David Banisar, deputy director of Privacy International, the combination of DoubleClick's over 100 million cookie-derived profiles with Abacus' database of purchasing habits could mean that the majority of web-connected Americans will lose their online anonymity [cited in WR00]. At the same time, data collected online can be used for targeting in offline media. Abacus touts its database is "now flagged with known Internet users", enabling its clients to identify and model Internet users in their existing address lists and targeting them separately in direct mail campaigns [AD00].

The NAI Principles: Opt-In and Opt-Out

Many users feel that having their names associated with detailed online and offline behaviour profiles is a violation of their privacy, and were glad to see the Federal Trade Commission (FTC) voice concern over DoubleClick's plans to merge PII and non-PII.

The FTC uses a set of information practice principles to evaluate online privacy issues [FTC00]: Firstly, users must be notified of profiling activities on web sites and be given the opportunity to decide whether they want to participate in those activities. Secondly, web sites must provide users with reasonable access to their individual data, and make reasonable efforts to protect that data from loss, misuse, alteration, destruction, or improper access. Consequently, the leading advertising agencies formed the Network Advertising Initiative (NAI) in 1999 with the goal to preempt government regulation of the online advertising market by defining a set of self-regulatory principles. In July 2000, they released a policy governing the use of consumer data for OPM. This document [NAI00] specifies that:

- Agencies are forbidden to use PII about *sensitive data* such as medical or financial records, sexual behaviour or orientation, or social security numbers.
- If *non-PII* is collected, publishing sites are required to post a privacy policy that states which information is used, and provide a way for users to deny the collection of data.
- The rules on *merging* PII with non-PII depend on which data is collected first:
    - If *PII is collected after the non-PII* (as in DoubleClick's case of acquiring the Abacus database), the agency may not merge the data sets without the user's prior consent.

- If *non-PII is collected after the PII* (e.g. through ongoing behaviour tracking after a user enters his personal information), the publishing site must notify the user of this fact before he is entering any information and provide him with a way to deny the collection of non-PII.

These rules employ two different models for getting a user's decision on allowing data collection: In the *opt-in* model, the user has to give the agency explicit permission to use his data – if he does not opt-in, the agency is not allowed to use the data. The *opt-out* model works the other way round: The agency may use the data as long as the user does not explicitly forbid it.

As a consequence of the NAI principles, DoubleClick cannot simply merge its database of online non-PII with Abacus' database of offline PII – it is only allowed to merge the profiles of users who explicitly opt-in. Additionally, DoubleClick provides users the opportunity to opt-out on its web site. This feature is achieved by simply storing a cookie with the value `OPT_OUT` instead of a unique identifier on the user's computer [DC00d].

Conclusion: Personalization vs. Privacy

The NAI principles and continuing FTC observation of the online advertising industry prevent large-scale loss of anonymity on the web today. The question remains how far personalization and profiling can go before privacy is violated. A definitive answer is hard to find since the detail of profiling, type of data, reach of network, and intent of personalization all are important factors.

DoubleClick believes that by targeting ads to to consumers, it actually makes advertising on the web less intrusive because users are not bombarded with repeat and irrelevant messages. However, there is no guarantee that targeted banners are more effective than other ads. In fact, click-through rates for banner ads have been dropping consistently since they were introduced in 1995. According to Ann Carey, director of interactive services at New York's Margeotes Fertitta & Partners agency, less than 0.5% of users click on a banner ad when they see it [cited in KK00].

This is one of the reasons why Yahoo, the web's most visited search portal [MT00], does not invest in consumer profiling technologies. Anil Singh, chief marketing officer at Yahoo says another reason is that the company does not want to risk destroying the trust people have in their brand: "We're totally unwilling to compromise that for a short-term advantage of getting more advertisers." [cited in JW00]

Bibliography

[AD00] Abacus Direct Corp.: Internet User Selects. Available online at http://www.abacus-direct.com/resource/factsheets/internetuser.asp (cited: November 2000)

[DC00a] DoubleClick, Inc.: Non-Personally-Identifiable Information. Available online at http://www.doubleclick.net/us/corporate/privacy/non-identify.asp (cited: November 2000)

[DC00b] DoubleClick, Inc.: Targeting and Reporting. Available online at http://www.doubleclick.net/us/publishers/ad-serving/dart/features/targeting.asp (cited: November 2000)

[DC00c] DoubleClick, Inc.: DoubleClick Privacy Statement. Available online at http://www.doubleclick.net/us/corporate/privacy/ (cited: November 2000)

[DC00d] DoubleClick, Inc.: Opt-Out. Available online at http://www.doubleclick.net/us/corporate/privacy/opt-out.asp (cited: November 2000)

[FGM99] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P. and Berners-Lee, T.: RFC2616: Hypertext Transfer Protocol – HTTP 1.1; June 1999. Available online at ftp://ftp.isi.edu/in-notes/rfc2616.txt (cited: November 2000)

[FTC00] Federal Trade Commission: Federal Trade Commission Issues Report on Online Profiling; July 27, 2000 press release. Available online at http://www.ftc.gov/opa/2000/07/onlineprofiling.htm (cited: November 2000)

[JW00] Weaver, Jane: Does DoubleClick track too closely?; in: ZDNet News, January 27, 2000. Available online at http://www.zdnet.com/zdnn/stories/news/0,4586,2428392,00.html (cited: November 2000)

[KK00] Kranhold, Kathryn: Web advertising not as hot as it seems; in: ZDNet News, June 27, 2000. Available online at http://www.zdnet.com/zdnn/stories/news/0,4586,2595172,00.html (cited: November 2000)

[MT00] Tchong, Michael: Search Engine Positioning; in: Iconocast, August 24, 2000. Available online at http://www.iconocast.com/issue/20000824.html (cited: November 2000)

[NAI00] Network Advertising Initiative: NAI Self-Regulatory Principles. Available online at http://www.networkadvertising.org/press/principles.pdf (cited: November 2000)

[NC96] Netscape Communications: Persistent Client State HTTP Cookies. Available online at http://home.netscape.com/newsref/std/cookie_spec.html (cited: November 2000)

[WR00] Rodger, Will: Activists charge DoubleClick double cross; in: USAToday.com, June 7, 2000. Available online at http://www.usatoday.com/life/cyber/tech/cth211.htm (cited: November 2000)