Web Tracking

SNET2 Seminar Paper - Summer Term 2011

Niklas Schmücker
Berlin University of Technology
Email: nschmuecker@gmail.com

***Abstract*—This paper gives an introduction to web tracking and provides an overview over relevant technologies currently found in use on the Internet. We examine motivations for web tracking and discuss issues related to privacy and security. Furthermore, we present and compare countermeasures intended to protect end users. We end with a discussion of possible future trends and developments in the field of user tracking.**

I. Introduction

Web tracking technologies are used to collect, store and connect user web browsing behavior records. The information gained thereby is of interest to various parties. Major motivations for web tracking are:

- Advertisement companies actively collect information about users and accumulate it in user profiles. These profiles are then used to tailor individualized advertisements. Instead of showing random ads to users, their profile information, for example, age, sex and other sites visited in the past, is taken to chose content of relevance to their interests. Thus, advertisers can focus their budget on consumers who are likely to be influenced. This use case is further discussed in Section IV.
- Law enforcement and intelligence agencies may use web tracking technologies to spy on individuals and to solve crimes. The unique identification of individuals on the Internet is important in the fight against identity theft and for the prevention of credit card fraud.
- Usability tests of web applications: By observing the steps an individual performs while trying to solve a certain task on a web page, usability problems can be discovered and fixed [1]. Details are found in Section III.
- *Web analytics*, a related field, focusses less on the individual user, but more on the performance of a web site as a whole. In the e-commerce business, shop operators use web analytics to maximize their revenue, for example, by evaluating which pages generate most income, which banner ads account for most traffic, or during which steps of the order process customers are lost. Web analytics solutions are further discussed in Section II.

Some of these use cases are more controversial than others. For example, usability tests with informed participants have few privacy implications, while tracking web usage and creating profiles of unsuspecting users for marketing purposes is arguable. Section VI discusses privacy protection options available to Internet users, while section VII gives a short overview over the current legal situation.

II. Web analytics

The web analytics field is concerned with the measurement and interpretation of web site usage data. A variety of information is potentially of interest to web site operators, such as:

- The number of visitors over time, which can further be divided into returning and new visitors. This includes how long individuals stay on the site and which pages they look at (also see Section V-A).
- How visitors find out about the web site. Usually three sources of traffic are differentiated: Direct traffic (the user enters the address into the address bar), traffic referred from other web sites (see Section V-C2), as well as search engine traffic. For the latter, even the relevant search keywords can be extracted.
- The effectiveness of marketing campaigns, measured by how much traffic the corresponding advertisements drive to the web site.
- The geographical location of visitors, which is usually inferred from their IP addresses[1].
- Company identification, which, for example, allows web masters to see which competitors are looking at their web site. This data is also based on IP address information.
- Technical details, such as operating system, screen resolution and web browser version of the visitors.

Web analytics software can be self-hosted, but more commonly third-party services, such as *Google Analytics*[2], are used. The collected data can usually be presented visually. Figure 1 shows an example of a web analytics dashboard screen.

These services usually require web masters to include a JavaScript code snippet into their web sites, which then downloads more tracking code from a third-party server. Whenever a user performs a certain action, such as transitioning to another page, the tracking code informs the analytics server of this event[3].

A different paradigm does not rely on client-side code, but instead extracts information directly from the web server's log files. An advantage of this approach is that it also works

[1]There exist several databases, both commercial and free, which map IP address ranges to geographical locations.

[2]http://google.com/analytics

[3]The Google Analytics code, for example, sends back data to the collection server by requesting a 1x1 pixel GIF image. The data is appended as a list of query parameters to the URL of the image GET request (see Section V-B8 for details).

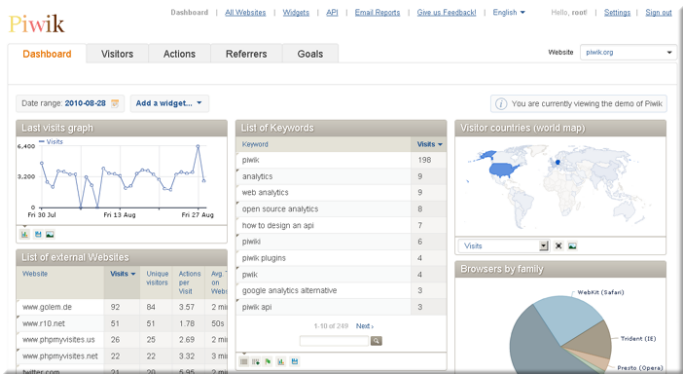

Fig. 1. The dashboard of *Piwik*, a free and open source web analytics software package (http://piwik.org). The displayed information includes geographical location, browser composition and number of visitors.

if the client has JavaScript turned off. However, client-side technology can collect more information on the local machine than the browser sends out by default.

III. Usability tests

Tracking technology can also be handy for usability evaluations of web applications, to help guide decisions in their design and development process. With JavaScript, it is possible to capture detailed records of user mouse and keyboard input [1]. Thus, the interaction with web sites can be analyzed in detail. For example, the following scenarios become feasible:

- To record and play back cursor movement paths at a later time, to see if users are having problems locating or using certain functionality on the web site. The movement data can be abstracted to a higher level, to give information about the interaction with certain page elements, such as buttons and scroll bars, instead of raw cursor coordinates.
- When users fill in questionnaires, the order in which they proceed and how long each step takes them can be used to guess which answers are causing problems and how the form fields could be rearranged.
- The data could even be used to classify the computer skills of the visitor and to adapt the page accordingly [1]. For example, more advanced options could be shown to experienced users, while someone struggling even with basic tasks would receive additional hints.

An advantage of in-browser, JavaScript-based tracking solutions is that they do not require special hard- or software. They make it possible to observe subjects in their natural home setting, which also lowers the costs of testing [1]. However, laboratories offer additional possibilities, such as using eye-tracking devices[4] to analyze which items on a page receive most attention.

IV. Internet marketing and advertising

The online advertisement market benefits strongly from web tracking technologies. This section gives a short overview of the field.

[4]https://secure.wikimedia.org/wikipedia/en/wiki/Eye_tracking

A. Contextual Advertising and Semantic Targeting

Contextual advertising refers to the presentation of ads that are related to the content of the web site. Typically, keywords representing the overall topic of the page are extracted from its raw text. Ads are then picked according to those keywords, thus being in direct correlation to the content. For example, a user visiting a web site about cars would ideally be shown car-related advertisements. This increases the chance of an ad catching catching the user's interest. Thus, advertisement companies can charge more money for their services.

Semantic targeting refines this concept by trying to identify the semantics of the displayed content, including the disambiguation of words with multiple possible meanings. This ideally reduces the number of misplaced ads[5][6].

Natural language *sentiment analysis* techniques can be used to infer the subjective attitude of the author towards the web content. If a certain topic is depicted in a negative light, an advertiser might choose not to have his ads displayed on the same site. The same is true for pages containing adult content, violence or offensive language.

A popular example of a contextual ad serving application is *Google AdSense*[7]. AdSense requires web site operators to insert a small snippet of JavaScript code into their pages, which then fetches and displays ads relevant to the content, based on high-value keywords extracted by Google bots. Advertisers can bid on these keywords through the related *Google AdWords* program[8]. The highest bids win and are shown to the viewer. In this scheme, an advertiser only pays if his ad is actually clicked.

Since contextual advertising is based on the contents of the web site alone, it does not depend on tracking techniques.

B. Behavioral targeting

Behavioral targeting or *behavioral advertising* is a form of targeted advertising which tries to guess appropriate ad content based on collected user profiles. These profiles may contain information such as sex, age group, location, estimated income and interests.

This allows advertisers to use their marketing budget more efficiently by only reaching people who are likely to become customers. It has been shown in studies that behavioral targeting significantly increases the effectiveness of online advertisement, making individuals more likely to buy an advertised product [2].

These profiles are, to a large extent, built from search queries and browsing history of the users. Someone who regularly visits football-related web sites, online car magazines and who has been observed shopping for men's fashion in the past is very likely to be male. This data can be augmented with public information from social networks such as Facebook, resulting in highly detailed profiles. Most social

[5]http://www.shmula.com/steve-irwins-death-contextual-advertising-gone-bad/194

[6]http://mashable.com/2008/06/19/contextual-advertising

[7]https://www.google.com/adsense

[8]http://adwords.google.com

networks simplify this process for the advertisers, some of them intentionally, for example, by including tracking code, which tells third parties which account on a social network belongs to which of their profiled users [3].

C. Compensation in online advertising

Various pricing models exist for online advertising. Common concepts are:

- *Pay-per-impression*: Advertisers pay depending on how often ads are displayed to visitors.
- *Pay-per-click*: Advertisers only pay when ads are actually clicked.
- *Pay-per-sale*: Ad publishers get a share of the order amount when a purchase is made through one of the advertisements. *Pay-per-action* is a more generic concept, where publishers get compensated for agreed-upon user actions, such as sign-ups for a newsletter.

User tracking techniques are required to enable affiliate concepts such as pay-per-sale, because the user following an advertisement on site A needs to be matched to the user making the purchase on site B.

If no preventive measures are taken, these systems can be exploited easily. For example, in a pay-per-click model, advertisements of a competitor might be clicked repeatedly by the same individual to exhaust the marketing budget of the former, a practice called *click fraud*. Thus, user identification methods are used to help discover and prevent suspicious patterns.

D. Advertising networks

Online advertising networks offer services that allow publishers, that is, web site operators, to sell advertising space to advertisers. Key functions and requirements of ad networks are:

- Aggregating ad content from paying advertisers and finding publishers who are willing to display that content.
- Providing server infrastructure where the ads are hosted.
- Making available software solutions which allow webmasters to easily integrate advertisement space into their existing site layouts.
- Honoring the preferences of advertisers, regarding where and how the ads should be displayed. For example, marketers might not want their company logo to be shown along with certain types of content. This requires contextual advertising algorithms (see Section IV-A).
- Tracking of Internet users across domains and creation of user profiles, to allow for behavioral targeting (see Section IV-B).
- Collection of statistics and creation of reports, for example, to give advertisers information about which ads are performing well. This helps advertisers to optimize their campaigns.

In practice, web masters usually have to include a small snippet of JavaScript code into their web sites, which then fetches and displays content from a third party ad server. Advertisers, on the other hand, usually create and track their marketing campaigns via a web interface.

Examples of popular ad networks are DoubleClick (owned by Google), ValueClick and AdBrite[9].

V. Web tracking technologies and concepts

This section discusses the technical background of the different forms of web tracking presented earlier.

A. Clickstream analysis

A *clickstream* is a recording of the actions performed by a user on a web site. *Clickstream analysis* deals with the collection and evaluation of this kind of data.

Conversion funnel analysis uses clickstream data to check where visitors enter the web application and then tracks their progress towards a certain predefined goal, such as the subscription to a newsletter or the completion of a purchase. This allows webmasters to see where in the process potential customers are lost.

Clickstream data has also been used, for example, to predict whether a user is likely to submit an order on an e-commerce web site [4], and to evaluate the effectiveness of banner advertising [5].

1) Gathering clickstream data: There are three main approaches that are used to collect click stream data.

1) The order in which pages have been requested by a certain user can be extracted from web server logs by the administrator of the corresponding web server.
2) Another approach uses client-side JavaScript tracking code in the browser that collects click data and sends it to a tracking server. Different users are typically distinguished with the help of cookies (see Section V-B2 for details). This approach is called *page tagging*. One advantage of page tagging over log file analysis is that only actual human visitors are counted and not web spiders and other types of robots. Page tagging also recognizes clicks which do not generate requests to the server, such as when a user accesses cached pages.
3) Internet service providers have full access to the traffic records of their customers and can gather clickstream data by *inspecting packets*, which is discussed in Section V-C1.

B. Client identification

The *Hypertext Transfer Protocol (HTTP)* is stateless by design[10], which means that different page requests are independent of each other. However, in todays web applications, it is often required to identify a user over many subsequent requests. For example, temporary session information has to be stored for online shops, where users add several products to their shopping carts before finally proceeding to the checkout. This is called *session handling*.

[9]Wikipedia has an extensive list of ad networks: https://secure.wikimedia.org/wikipedia/en/wiki/List_of_advertising_networks

[10]See the corresponding RFC: http://www.w3.org/Protocols/rfc2616/rfc2616.html.

As such, session handling has similar requirements to user tracking: Page requests have to be matched to individuals, or rather to their browsers running on particular computers. There are various approaches to accomplish this form of user differentiation. The simplest one is based on IP addresses. More powerful techniques assign a unique identifier to each individual and session, which is initially stored somewhere on the client and subsequently transmitted to the server with every request. The following sections go into detail about these and similar techniques.

1) Identification by IP Address: Every computer on the Internet has an Internet Protocol address, which is used to route data towards it. For various reasons, using this address alone to match a request to a user does not work in todays Internet:

- A single public IP address is often shared between many endpoints, e.g., because of NATs[11] and proxies (see Section VI-A1).
- IP addresses can be faked, which introduces security problems.
- Typically, end users with dial-up connections change their IP addresses frequently.
- A single user may have multiple parallel sessions open.

Hence, assisting techniques are required in most cases.

2) HTTP Cookies: HTTP cookies are arbitrary name-value pairs stored in the web browser. Cookies are commonly used for session handling, storage of site preferences, authentication and the identification of clients. Typically the short form "cookies" refers to HTTP cookies, even though similar mechanisms exist which are referred to by the same name.

When a web browser initiates a request to a server, the latter can ask for the instantiation of a cookie on the client in its response. If accepted by the browser, the cookie is stored and sent to the server inside a HTTP header field whenever the client makes a new request to the same domain.

Often session IDs, unique random numbers used to identify a session, are stored inside cookies, such that servers can match requests to a particular browser they have seen previously, without relying on IP address information. Due to a mechanism called the *same-origin security policy*, cookies are not sent to other domains, as they may contain confidential information.

Third party cookies are not set for the domain the user is currently viewing, but for external domains from which additional data, such as images and scripts, was fetched. Third party cookies are sent to the corresponding server no matter which page the user is currently viewing, as long as it includes content from said third party.

HTTP cookies without an expiration date are automatically deleted when the browser is closed. However, expiration dates can be many years into the future.

3) Adobe Flash Local Shared Objects: Adobe Flash is a popular browser plugin which is mainly used for animated and interactive web content. By default, the Flash browser plugin allows servers to store *Adobe Flash Local Shared Objects*, also called *Flash Cookies*, which are similar to HTTP cookies and can be used for the same purposes. However, they are managed by the Flash plugin itself and not by the web browser.

LSOs were introduced by Adobe to get around restrictions of traditional cookies [6]. By default, each web site can store up to 100 KB of cookie data, while a single HTTP cookie is limited to 4 KB. Persistent Local Shared Objects do not require an expiration date and can only be deleted easily in very recent browsers that use Adobe's *NPAPI ClearSiteData API*[12], which only became available in 2011[13].

For older browsers, the user has to install special browser plugins[14] or use a tool only accessible on Adobe's web site.

Adobe Flash includes methods for developers to bypass the same-origin policy, which normally prevents sites from accessing data stored by other domains. In many cases Flash does not warn or prompt the user when content is written or accessed [7].

Using executable Flash code, LSOs can be enriched with specific personal and technical information, such as user name, computer name and files on disk. LSOs are shared between all applications on the computer using the Flash plugin, thereby making it possible to identify users across multiple browsers.

4) Web storage: Web storage is a specification by the *World Wide Web Consortium*[15], which addresses storing and accessing big chunks of data and key-value-pairs in a web browser via client-side scripting.

Web storage is supported by all major browser vendors [8]. The specification includes two JavaScript objects, *localStorage* and *sessionStorage*. The former is used for persistent storage of data, while the latter is cleared on browser termination. For example, the JavaScript code *window.localStorage.setItem("userid", 12345)* sets a persistent user ID, and *window.localStorage.getItem("userid")* is used to access it again at a later point in time. Unlike cookies, the content of the web storage is not sent to the server with every request, but has to be transmitted explicitly using JavaScript code.

5) Silverlight Isolated Storage: Silverlight Isolated Storage is similar to web storage and can be used to store data locally on the user's computer, such as key/value pairs and arbitrary files [9]. However, it requires the user to have the Microsoft Silverlight plugin installed.

6) Google Gears: Gears is an architecture by Google that allows web sites to save data locally such that basic functionality can also be accessed without being connected to the Internet. The user has to give explicit permission for every site that wants to access the store. In Gears, data is

[11] With the ongoing IPv4 address exhaustion, NATs will probably become even more prevalent until IPv6 finds widespread deployment.

[12] https://wiki.mozilla.org/NPAPI:ClearSiteData

[13] http://blog.chromium.org/2011/04/providing-transparency-and-controls-for.html

[14] *BetterPrivacy* (http://netticat.ath.cx/BetterPrivacy/BetterPrivacy.htm) is specifically designed to clean LSOs and similar unwanted information that is else hard to get rid of.

[15] http://www.w3c.org

not shared between different browsers on the same computer. A strict same-origin policy is followed, which prevents sites from reading arbitrary saved content. Hence, Gears poses less problems from a privacy standpoint than other mechanisms [7].

7) Hidden Form Fields: The *Hypertext Markup Language* includes *form* elements, which are intended to allow users to input data for subsequent transmission to the server via the POST or GET methods.

HTML forms can, however, also be used for session management: A form, hidden from the view of the user, stores a unique session identifier as its value. When the transition to another page occurs, the value is automatically sent to the server as a POST or GET, which allows the server to match the request to a session. The response can then be tailored accordingly and typically includes a hidden form field prefilled with the same session information. This principle is similar to cookies, even though relying on a different technical implementation. However, the data is not persisted to disk.

8) URL Query Strings: Query strings are pieces of information appended to the end of URLs, which are sent to the server when the corresponding link is accessed.

For the purpose of user identification, the web server appends an identifier to the links held by the web site. For example, the URL *http://www.shop.com/somepage.html?sessionid=123* includes the session ID "123". When such a link is followed, the browser sends the query parameter as part of its HTTP GET request to the server, which can then parse it and use it to identify the session. Subsequent pages delivered to the user will carry the same identifier appended to the links.

A disadvantage of this approach is that sensitive session information is included in the URL, which is leaked then the latter is shared. Hence, web applications should not rely on query strings alone [7]. Nevertheless, query strings are sometimes used as a fallback mechanism in case cookies are not available. Similarly to Hidden Form Fields, the data is not persistent.

9) HTTP authentication: HTTP natively supports authentication mechanisms, such as *Basic access authentication* and *Digest access authentication*. When accessing a web page with authentication turned on, the browser prompts the user for credentials and stores them temporarily. For every subsequent request, these credentials are submitted to the server within the HTTP *authorization header*, which can be used to identify the session and with it the user.

10) window.name DOM Property: The *Document Object Model*[16] of common web browsers includes the property *window.name*. It is accessible via client-side JavaScript code and can typically store several megabytes of data. Each browser tab has its own window.name property, which is empty just after creation. However, all pages accessed via links in a tab share the same window.name field, meaning that it can be used to exchange information between domains, which poses security and privacy threats. For example, the sequence of visited pages could be stored.

[16]http://www.w3.org/DOM

C. User tracking technologies

User tracking has similarities to session handling. Both require the clear identification of a client machine. Hence, techniques from session handling, such as cookies, can also be used for tracking. An overview is given in the following sections.

1) Deep packet inspection: Internet service providers have full access to the traffic data of their customers. Deep packet inspection refers to the practice of not only looking at IP packet headers to determine source and destination of a message, but also analyzing the actual payload.

If no encryption is used, this allows ISPs to see exactly what their customers are doing on the Internet, not limited to web browsing activities. The structure of a generic IP version 4 packet is shown in Figure 2.

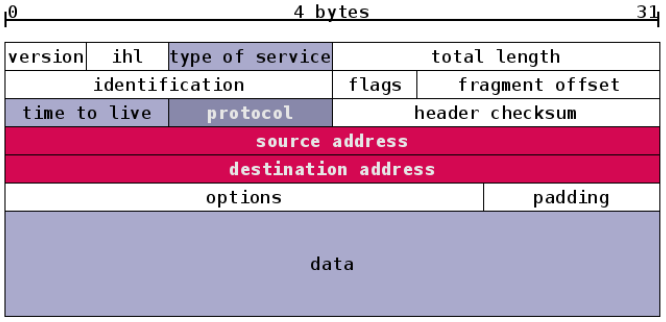

Fig. 2. The structure of a generic IPv4 packet. While traditional traffic analysis mechanisms only look at the header information, such as source and destination address, deep packet inspection also considers the actual data payload. Taken from http://www.postfixvirtual.net/dnsbind.html.

2) HTTP referrer: The *HTTP referrer* is a field in the HTTP header which contains the page that the user originated from. Using HTTP referrers alone, it is not possible to build extended user tracks across arbitrary domains, but often revealing one step back in the browsing history is already more than the user is comfortable with. For example, if a user enters a web site from a URL identifying a public user page in a social network, the latter can be linked on tracking servers to existing user records, thereby creating a rich profile. It may not be immediately clear that the referring account page really belongs to the visitor himself. However, in some cases the referrer URL is associated with profile editing or similar actions that can only be done by the owner of a profile [3].

The browser has full control over hte referrer header it sends. As a countermeasure, the HTTP referrer can be set to an arbitrary value or even left blank[17].

The HTTP referrer is extensively used in web analytics, because it shows which sources and marketing campaigns drive traffic to a web site.

[17]RefControl is a browser plugin for Firefox that forges the referrer: http://www.stardrifter.org/refcontrol.

3) Web bugs and tracking cookies: As already mentioned, cookies are not only useful for session handling, but can also be used for doubtful user tracking practices.

For the creation of extensive user profiles, users must be identified and matched across multiple web sites. First party cookies are not sufficient, because they are only made available to the server corresponding to the domain of the web site the user is currently viewing.

Third party cookies (see Section V-B2) are used to work around this restriction: Services wishing to track users can employ invisible dummy images[18], so called *web bugs*. When a web site containing a web bug is accessed, a request is made to the domain hosting the web bug. This has two implications:

- The third party tracking service knows when a certain IP address has accessed the web site containing a specific web bug.
- The third party can set a cookie containing a unique ID on the client machine, a so called *tracking cookie*. This cookie is sent back to the third party whenever the user views a page which includes one of its web bugs.

Hence, if a third party places web bugs on multiple web sites, a browser can be tracked across domain boundaries. Some browsers can be configured to only accept third-party cookies if the corresponding third-party server declares its intended use of the collected data[19], and almost all browsers can be configured to reject cookies from third parties entirely. However, this usually only prevents their creation, while existing ones are readily sent to the foreign server. Hence, a common trick is to redirect users to a page belonging to the tracking network, which then sets a first-party cookie. This cookie can later be read by web bugs on other sites.

The controversial Facebook *Like button* can be seen as a web bug. It is typically included as an inline frame (*iframe*), that is, an own HTML page nested inside the current web site. When this inline page is requested, the address of the main page, along with Facebook's session cookie, is sent to the Facebook servers. This allows Facebook to see which other pages their customers are browsing [10] [11].

Some web analytics services, such as Google Analytics, only use first-party cookies. A technique called *cookie handover* makes it possible to track users across domains anyway: JavaScript code on domain A appends a query parameter containing the current user ID, read from the cookie, to all outgoing links. When a link leading to domain B is clicked, JavaScript code on domain B parses the query parameter and creates its own first-party cookie with the same user ID. Thus, the cookie gets cloned. This approach only works if the prepared links are used to access domain B.

4) Zombie cookies: While cookies are in principle an effective mechanism to track users, more people are becoming aware of the privacy implications and clear their cookies regularly [12]. While the removal of HTTP and Flash cookies can be achieved without much effort, *Zombie cookies*, also called *Super-cookies*, are designed to be resist deletion efforts [7].

Companies in the field of Internet marketing and advertising, such as Google [13], are particularly interested in this technology, because it prevents users from easily changing their online identity. For example, the removal of cookies may lead to the same person being counted multiple times by web analytics software, which causes fragmentation in the recorded data. This has unwanted effects. For example, an advertiser using banner ads may not want to pay more than once when the same individual clicks a banner multiple times.

Zombie cookies store the identifying user information redundantly in many places. These locations include:

- HTTP cookies,
- Flash cookies,
- Silverlight Isolated Storage,
- Web storage,
- Web history,
- browser cache,
- the window.name DOM property.

Whenever one of these stores is cleared, the Zombie cookie uses client-side JavaScript code to recreate it from the remaining data [12]. Hence, the removal of Zombie cookies is tedious and requires substantial effort [14].

Zombie cookies that use stores which are accessible from several applications, such as Adobe Local Shared Objects (see Section V-B3), can spread across different browsers on the same computer. A survey has found that some of the top 100 web sites are actively using Zombie cookie mechanisms to recreate deleted user identifiers [12]. Evercookie[20] is a well-featured and popular open source Zombie cookie implementation [15], even though proprietary solutions may include more advanced techniques such as browser fingerprinting.

5) Browser Fingerprinting: The *Electronic Frontier Foundation*[21], an US-based civil liberties group, has recently demonstrated the feasibility of a novel approach to browser identification, called *browser fingerprinting* [16].

During browser fingerprinting, seemingly insignificant and non-critical configuration and version data is collected from the web browser, for example:

- The browser's *user agent information*[22],
- the client's screen resolution,
- the local timezone,
- the list of installed browser plug-ins,
- the list of installed system fonts,
- the operating system,
- the browser's language,
- the list of accepted MIME types.

Some of this data can be inferred from the HTTP headers alone. A typical HTTP request header is shown in Figure

[18] E.g., 1x1 pixel images or transparent images. In fact, web bugs are not restricted to images. They can be in any file format which can be embedded into web pages.

[19] This is usually communicated through the *P3P* protocol, see http://www.w3.org/P3P.

[20] http://samy.pl/evercookie

[21] https://www.eff.org

[22] https://secure.wikimedia.org/wikipedia/en/wiki/User_agent

3. However, other parts must be harvested using client-side JavaScript, which then sends this information back to the server. All of this may happen without the knowledge and consent of the user [16].

Each single piece of information, even though not of much use on its own, reduces the entropy of the identity of a browser to a certain extent[23]. If enough information adds up, the browser can be identified uniquely.

Request parameter	Value
Requested URI	/headers
Request Method	GET
Remote IP Address	85.177.91.142
Remote IP Port	34561
Protocol version	HTTP/1.1
HTTP Header*	**Value**
Accept	text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Charset	ISO-8859-1,utf-8;q=0.7,*;q=0.7
Accept-Encoding	gzip, deflate
Accept-Language	en-us,en;q=0.5
Connection	keep-alive
Dnt	1
Host	www.xhaus.com
Keep-Alive	115
User-Agent	Mozilla/5.0 (X11; Linux x86_64; rv:2.0.1) Gecko/20100101 Firefox/4.0.1

Fig. 3. Typical HTTP request headers, including information such as the user-agent string and the accepted MIME types.

Eckersley notes that much more information can be collected using browser plug-ins such as Java, Silverlight and Flash, hence adding even more information to the fingerprints [16]. Commercial fingerprinting services include these and more elaborate techniques, such as measuring processor speed, the time difference between the clocks on the client and the server side, as well as properties of the client TCP/IP stack [16].

Using certain supposedly privacy-enhancing software can have paradoxical effects. For example, when a Flash blocker is active, this prevents the fingerprinting algorithm from obtaining a list of system fonts. However, the Flash plugin is still being detected. This abnormality can make the browser fingerprint more revealing than without a blocker. The same can happen for browsers using faked user agent strings. If the rest of the configuration does not match what is reported otherwise, this can result in a unique browser fingerprint.

In the ideal[24] case, this signature alone can be used as a global identifier for a browser, thus rendering other identification methods unnecessary. When this is not the case, that is if the fingerprint on its own is not unique, it can still be combined with other information, such as the user's IP address. Fingerprints can also be used by Zombie cookie implementations to aid the automatic recreation of deleted user identifiers (see Section V-C4).

[23]An explanation of entropy in the context of browser identification can be found at https://www.eff.org/deeplinks/2010/01/primer-information-theory-and-privacy

[24]Or worst, from a privacy standpoint.

While the fingerprint usually changes even after small modifications to the browser configuration, such as updates of plug-ins or modifications of the screen resolution, Eckersley reports that using a simple heuristic algorithm, newly appearing fingerprints could be matched to earlier ones with an accuracy of about 99% [16].

Fingerprinting can be exploited by companies in a similar fashion to cookie- and IP-address-based tracking. It does not leave persistent evidence behind on the client computer. In an experiment[25] conducted with almost a million visitors, the EFF found that about 84% of visitors had unique browser fingerprints. For browsers with certain plugins installed, such as Flash and Java, the percentage was even higher with about 94% uniquely identified browsers, even though the authors note that their sample of visitors may be biased.

VI. Privacy-enhancing Technologies

Certain countermeasures exist that help individuals to protect themselves from being identified and tracked on the Internet.

A. Hiding IP addresses

The Internet Protocol address can be used to identify endpoints. Administrators can see the IP addresses of visitors in their server log files. Internet service providers even have means to perform *traffic analysis*, where the parties communicating with each other on the network are determined, by inspecting the source and destination fields in IP packet headers. Deep packet inspection techniques go even further by also scanning the actual data payload in the communication flow (as discussed in Section V-C1). Some ISPs have admitted to selling these customer Internet browsing records to marketers [17].

ISPs can match IP addresses directly to paying customers registered with them. Even though identification of individuals on the basis of IP addresses alone is not easily possible for web site administrators, IP addresses can be combined with other identifying data. Thus, users with privacy in mind have benefits from hiding their IP address, for which several solutions exist.

1) Proxy servers: Proxy servers, also called proxies, act as intermediaries that forward requests on behalf of clients. All data, including responses from the destination, flow through the proxy, such that the IP address of the client computer is only revealed to the proxy. The destination server only sees the address of the latter.

Besides anonymization purposes, proxy servers are used for a variety of reasons, for example, caching, content filtering, logging, data leak prevention and malware scanning. End users have to trust the proxy server, as its administrator and other individuals who have gained control over the proxy server can inspect all traffic, which means that browsing behavior and even passwords can be logged. Thus, it is not advised to send unencrypted data over proxies that can not be trusted.

However, even if the traffic is encrypted, the IP headers are not, so the two communicating endpoints can still be identified

[25]https://panopticlick.eff.org

by someone observing the network. A user can chain several proxies, in which case it becomes harder to reconstruct the original source. *Virtual Private Networks*[26] can be used in a similar fashion to proxies. Several providers offer proxy- and VPN-based anonymization services.

2) Tor and Privoxy: Tor is a network of virtual tunnels originally developed for the purpose of protecting government communications [18] [19]. Today it is open to the public and serves as one of the major anonymization infrastructures. The Tor network is operated by volunteers around the world. Tor hides the IP addresses of end users, while working around some of the weaknesses of traditional proxies.

An *onion routing* mechanism [20] is used, where encrypted connections between multiple chained relays ("Tor nodes") are established. Applications do not contact the destination server directly, but instead send data through a *Tor proxy*, which serves as the first node in the chain. It is typically run on the local machine and can be accessed by any TCP application with *SOCKS* support[27], such as web browsers, IRC clients and instant messengers.

The Tor proxy picks a path through the Tor network, which includes at least three different Tor nodes[28]. Secured by public key cryptography[29], symmetric encryption keys are then negotiated between the Tor proxy and each of the nodes.

The symmetric keys are used by the Tor proxy to wrap the actual data packets inside multiple layers of encryption[30]. The packets are then forwarded to the first hop on the path, which removes the outermost encryption layer and extracts the address of the next node in the chain, to which the packages are passed on. This process is repeated until the final destination is reached.

Thus, the intermediary nodes do not know the source, destination and contents of the original packets. No individual relay node can deduce the complete path a packet has taken. While the first node in the circuit can identify the sender, it does not know the contents of the message, nor its destination. A path through the Tor network is only valid for a certain timeframe, after which a new one is chosen. This makes it harder to link earlier actions to new ones.

Recent versions of Tor also include an own Domain Name System resolving mechanism that prevents the leakage of DNS requests, which could give away what web sites the user is trying to connect to.

However, Tor is not secured against all kinds of attacks. For example, end-to-end timing methods, where statistical analysis is used to correlate traffic originating from one computer with traffic arriving at another computer [21].

[26]https://secure.wikimedia.org/wikipedia/en/wiki/Virtual_private_network

[27]https://secure.wikimedia.org/wikipedia/en/wiki/SOCKS_(protocol)

[28]Tor could in principle be forced to use less nodes. However, this would allow a single malicious node to learn about the identity of the user. See https://gitweb.torproject.org/torspec.git/blob/HEAD:/control-spec.txt for details.

[29]Wikipedia has a good explanation of public key cryptography: https://secure.wikimedia.org/wikipedia/en/wiki/Public-key_cryptography.

[30]Using public key cryptography for the data transfer would be a lot more computationally demanding.

Other research has shown that fingerprinting-based methods can be used by attackers on the same network as the Tor user to predict what pages the latter is viewing [22].

Tor does not encrypt traffic from end to end: The final node in the chain, also called *exit node*, removes the last layer of encryption and passes the packets to the destination, as visualized in Figure 4. Thus, both the administrator of the exit node, as well as individuals with access to the network between the exit node and the destination, can eavesdrop on the traffic, in case no additional end-to-end encryption is used.

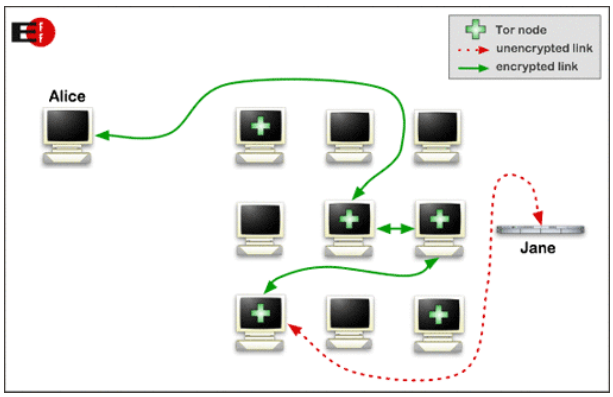

Fig. 4. *Tor routing*: Alice picks a random path through the Tor network. The "last mile" between the Tor exit node and Jane is not encrypted. Taken from https://www.torproject.org/about/overview.html.en.

Like traditional proxies, Tor does not prevent non-IP-address-based types of tracking. Hence, it is often used in conjunction with other software, such as *Privoxy*[31]. Privoxy is put in front of the Tor proxy. It filters advertisements and enhances privacy by stripping outbound information contained in HTTP headers and cookies, thus preventing the browser from leaking data to the server. Because the Privoxy server can be accessed just like a normal Proxy, it works with all browsers.

B. End-to-end encryption

Deep packet inspection (see Section V-C1) by entities with control over the network can be defeated with the help of *end-to-end encryption*. For example, the *HTTPS* protocol, a combination of *HTTP* and *TLS*, is commonly used to protect web traffic. Similar to Tor, HTTPS uses the public key of the web server, published in its certificate, to negotiate a symmetric encryption key, which is then used to secure all further communication.

C. Browser-based blocking and plugins

Most browsers provide ways to prevent users from being tracked, either directly or via browser extensions. Some more popular examples are discussed in the following sections. However, end users need to be suspicious when installing new browser extensions from unknown sources, as there have been cases of malicious extensions gathering browsing data without user consent [23].

[31]http://www.privoxy.org

1) Tracking Protection Lists: Microsoft Internet Explorer 9 has out of the box support for *Tracking Protection Lists* [24]. These lists define domains from which content is only fetched if entered into the address bar or directly clicked on by the user, thus preventing cross-browser tracking from third-party domains.

2) Torbutton: Torbutton[32] is a Firefox extension which lets users enable or disable Tor in the browser with one click. Apart from changing the browser's proxy settings, it provides additional privacy enhancing functionality, some of which is listed below.

- Can restrict JavaScript code.
- Resizes window dimensions to popular values to make fingerprinting harder.
- Blocks browser history reads, which prevents web sites from knowing which other sites a user has visited.
- Clears HTTP cookies and DOM storage.
- Makes the user agent string more generic, for example, by setting the reported language to English.
- Prevents the browser from writing cache to disk, as it could include unique tracking identifiers.
- Disables plugins which could work around the browser's proxy settings.
- Controls the HTTP referrer header.

Eckersley reports that Torbutton in conjunction with Tor is an effective countermeasure against common forms of browser fingerprinting [16].

3) Adblock Plus: Adblock Plus is a browser extension which supports *Mozilla Firefox* and *Google Chrome*. Its main functionality is blacklist-based hiding of advertisements. The user chooses from a number of subscription lists, which are regular expression based filters that stop unwanted content embedded into web pages from being downloaded. The most visible benefit to the user is the reduced number of ads.

EasyPrivacy[33] is a filter subscription for Adblock Plus specifically designed to protect the privacy of its users. It stops tracking content, such as web bugs and unwanted JavaScript files, from being downloaded. Thus, requests for web bugs never reach the tracking servers.

4) NoScript: NoScript[34] is a browser extension available for *Mozilla Firefox* that selectively blocks JavaScript, Java, Silverlight, Flash and other executable content. The standard behavior is *default deny*, thus only allowing content that the user has explicitly given permission to. While NoScript is mainly used for security reasons, it also disables web tracking services that rely on active client-side content.

Since many modern web sites use scripts and plug-ins for legitimate reasons, the whitelist-based approach implemented by NoScript has usability drawbacks and requires frequent user intervention.

5) RequestPolicy: RequestPolicy[35] is a browser extension for Mozilla Firefox which gives the user control over which other domains a web site is allowed to make connections to. Like NoScript, it follows a whitelist-based approach, meaning that initially all *cross-site requests* are forbidden. An example of the user interface is shown in Figure 5.

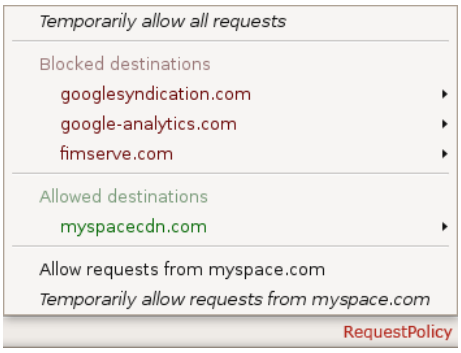

Fig. 5. *RequestPolicy* gives fine-grained control over which other domains a web site is allowed to make connections to.

Cross-site requests include any request a browser makes from the current site to a third-party domain. Examples are downloads of external content, such as image, CSS and JavaScript files, but also redirections to other web sites initiated by JavaScript or *META refresh tags*[36], as well as browser prefetch instructions[37], which are normally used to speed up the browsing experience.

By blocking requests to advertiser networks, RequestPolicy protects the privacy of the user. Besides, it also offers several security benefits, such as preventing *Cross-Site Request Forgery*[38] attacks, where a site sends requests to another web site, making it look as if they where coming directly from the user.

Similar to NoScript, a major usability drawback of RequestPolicy is its strict *default deny* ruleset, which breaks most modern web sites and requires active participation of the user.

D. Private Browsing Modes

All major browser vendors now include *private browsing* modes into their browsers, under various names. In Safari and Firefox, this feature is called "Private Browsing", in Chrome "Incognito mode", in Opera "Private Tab/Window" and in Internet Explorer "InPrivate Browsing".

In private mode, typically cookies and other browser persistence mechanisms are disabled. No browser history is recorded, and writing of caching information to disk is prevented. Usually passwords and contents of form fields are not saved. The intention of private browsing modes is to allow users to temporarily put down their normal online identity and visit particular web sites without leaving local traces. Ideally, a browser in private mode contains no personal identifying information and looks like a clean install to a web site.

However, private browsing modes are primarily designed to prevent individuals with access to the local computer to

[32] https://www.torproject.org/torbutton
[33] https://easylist-downloads.adblockplus.org/easyprivacy.txt
[34] http://noscript.net
[35] https://www.requestpolicy.com
[36] https://secure.wikimedia.org/wikipedia/en/wiki/Meta_refresh
[37] https://developer.mozilla.org/En/Link_prefetching_FAQ
[38] https://secure.wikimedia.org/wikipedia/en/wiki/Csrf

see what previous users did on the web. As such, no explicit attempts are made to disguise the user's IP address, or to fool fingerprinting services. Hence, sophisticated tracking services might still be able to identify users uniquely. Also, McKinley notes that, as of 2010, no browser managed to provide a private browsing implementation that prevented all types of information leakage between private and normal browsing modes [7].

E. Opt-out cookies

Some companies that perform web tracking offer users the possibility to opt out of their program by setting a special opt-out cookie, which gets recognized and honored by their tracking servers[39]. *Google* and the *Network Advertising Initiative*, an cooperative of online marketing and analytics companies, are proponents of this approach [25].

Opt-out cookies have usability disadvantages. Users who care about their privacy and regularly clear their cookies may also accidentally delete their opt-out cookies. Furthermore, the creation of opt-out cookies requires the user to know about and to find the web site of each advertisement network they want to opt out from.

However, browser extensions such as *Beef Taco*[40] exist which help users with the management of opt-out cookies.

Another weakness of the opt-out approach is the disagreement about what an opt-out request implies. Some companies will still record profile data and only refrain from using it for targeted advertisements. Opt-out cookies require the full cooperation of the tracking industry.

F. HTTP Do not track header

A similar opt-out approach is the *HTTP do-not-track header*. When set by the browser, it signalizes to the web server that the user does not want to be tracked. This technology has not been standardized yet. A standardization draft proposal has been submitted to the *Internet Engineering Task Force* in March 2011 [26]. No modification to the HTTP standard is required, as the IETF RFC 2616 allows for custom HTTP headers.

Recent versions of popular browsers, such as *Mozilla Firefox 4*, *Microsoft Internet Explorer 9* and *Apple Safari 5.1* already implement do-not-track headers [26]. In older browsers, support can be added via browser extensions. For example, recent versions of Adblock Plus and NoScript implement the do-not-track header.

However, similar to opt-out cookies, servers are in no way forced to honor the do-not-track request. Because the header has only been introduced very recently and because there are no laws concerning it yet, most of todays web servers will simply ignore it. A regulatory framework with effective enforcement mechanisms is still required. But even with laws in place, the do-not-track header should best be used in combination with other techniques.

The main advantage of the do-not-track header over opt-out cookies is that it represents a single persistent setting that works across all web sites, while opt-out cookies need to be set for each advertising network individually.

VII. Legal Situation and Proposed Bills

In the European Union, a privacy directive applicable to all member countries[41] forces web sites to inform users what data is stored on their local computer and for what purpose. Web sites have to refrain from such practices if the user does not agree. A study has shown that this directive has significantly reduced the effectiveness of online advertisements on European sites [2]. In November 2009, this EU directive has been revised and now imposes even more restrictions, for example, regarding the use of Flash cookies. However, the revision has not been implemented in the national laws of all member states yet [27].

In Germany, the Bundesdatenschutzgesetz states that collection and retention of personal data is only permitted if explicitly allowed by statutory provision, or if user consent is obtained in advance. According to the Telemediengesetz, service providers may only store data without consent if required for billing purposes. Hence, in Germany, anonymous user tracking for marketing purposes is only legal if the user is notified in advance and given the option to opt-out. Recording of personally identifying information for marketing purposes, such as IP addresses, is only allowed after explicit opt-in [28].

In contrast, such stringent regulations do not exist in the United States yet. The U.S. Federal Trade Commission is a major driver of consumer protection, but until now has only given recommendations. The FTC itself does not have authority to regulate user tracking and behavioral advertising.

However, a so called "Do-Not-Track Online Act of 2011" has recently been introduced by Senator John Rockefeller[42], built on recommendations by the FTC[43]. Congresswoman Jackie Speier has proposed a similar bill, called the "Do Not Track Me Online Act of 2011."[44] These two bills, if accepted, would force companies to honor the do-not-track headers sent by browsers, requiring the destruction or anonymization of any user information that is no longer immediately required for conducting business.

The "The Commercial Privacy Bill Of Rights"[45], proposed by Senators John Kerry and John McCain, has similar intentions, but, against the recommendations of the FTC, does not include the do-not-track mechanism. The bill would require

[39] The NAI offers consumers to opt out of various web tracking systems under the following URL: http://www.networkadvertising.org/managing/opt_out.asp.

[40] https://addons.mozilla.org/en-us/firefox/addon/beef-taco-targeted-advertising

[41] http://europa.eu/eur-lex/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf

[42] http://rockefeller.senate.gov/

[43] http://www.washingtonpost.com/blogs/post-tech/post/sen-rockefeller-to-introduce-do-not-track-bill/2011/05/06/AFphJN8F_blog.html

[44] http://speier.house.gov/uploads/\Do%20Not%20Track%20Me%20Online%20Act.pdf

[45] http://kerry.senate.gov/work/issues/issue/?id=74638d00-002c-4f5e-9709-1cb51c6759e6

companies to implement security measures to protect their customer's information, give customers the option to opt-out of data collection, and even require opt-in for the collection of specific critical data. In addition, companies would have to ensure that third-parties that they share consumer data with follow the same standards.

Several big companies, such as Google, Yahoo and Facebook, lobby against these bills. In their opinion the Internet as it is known today, offering many free, ad-financed services, could no longer exist [29].

VIII. Conclusion

Various techniques that allow for user tracking on the Internet and the motivations behind them have been examined. Regulation of the use of these technologies for marketing purposes protects consumers who are troubled by the collection of personal information, consumers who are not even aware that data is being collected, as well as consumers who do not see the privacy implications yet. On the other hand, regulation makes it harder for companies to innovate and market their products effectively. Privacy regulations that decrease the effectiveness of ads may lead to the decline of free content on the Internet. Hence, lawmakers must find a trade-off between the different interests.

On the other hand, with smart phones featuring geographical positioning technology becoming more and more popular, new issues have to be addressed. This has recently been shown when it became public that both Apple iPhone and Android devices have been storing location data without user consent [30] [31].

Such data is of interest to marketers, as it allows for the creation of real life customer behavior profiles. Location-based social networks, such as *Foursquare*, are already using location information for targeted advertising [32]. Online profiles and movement profiles combined mean an even deeper invasion of privacy. Hence, the *EU Data Protection Working Party*[46] is discussing possible regulations, such as requiring applications to inform their users when location information is used and to limit the time such data is stored [33][47].

References

[1] R. Atterer and M. Wnuk, "Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction," *Proceedings of the 15th International World Wide Web Conference Committee*, 2006. [Online]. Available: http://portal.acm.org/citation.cfm?id=1135811

[2] A. Goldfarb and C. Tucker, "Online advertising, behavioral targeting, and privacy," *Communications of the ACM*, vol. 54, no. 5, pp. 25–27, 2011. [Online]. Available: http://portal.acm.org/citation.cfm?id=1941498

[3] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," *Proceedings of the 2nd ACM workshop on Online social networks*, p. 7, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1592665.1592668

[4] W. Buckinx and D. Van Den Poel, "Predicting Online Purchasing Behavior," *Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium*, 2003.

[5] D. Hoffmann, T. Novak, and P. Chatterjee, "Modeling the clickstream: Implication for web-based advertising efforts," *MARKETING SCIENCE*, vol. 22, no. 4, pp. 520–541, 1998.

[6] Adobe, "Adobe Flash Player : What Is a Local Shared Object?" [Online]. Available: http://www.adobe.com/products/flashplayer/articles/lso/

[7] K. McKinley, "Cleaning Up After Cookies," 2010. [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Cleaning+Up+After+Cookies+Version+1.0#0

[8] World Wide Web Consortium, "Web Storage." [Online]. Available: http://www.w3.org/TR/webstorage/

[9] Microsoft, "Microsoft Silverlight Isolated Storage." [Online]. Available: http://www.silverlight.net/learn/quickstarts/isolatedstorage/

[10] J. Schmidt, "Das Like-Problem," *Heise Security*. [Online]. Available: http://www.heise.de/security/artikel/Das-verraet-Facebooks-Like-Button-1230906.html

[11] A. Efrati, "Like Button Follows Users," 2011. [Online]. Available: http://finance.yahoo.com/family-home/article/112769/like-button-follows-users-wsj

[12] A. Soltani, S. Canty, Q. Mayo, L. Thomas, and C. Hoofnagle, "Flash cookies and privacy," *SSRN preprint (August 2009)*, pp. 1–8, 2009. [Online]. Available: http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/download/1070/1505

[13] Google, "Cookies - Google Analytics." [Online]. Available: https://code.google.com/apis/analytics/docs/concepts/gaConceptsCookies.html

[14] J. TIMMER, "It is possible to kill the evercookie." *Ars Technica*, October 2010. [Online]. Available: http://arstechnica.com/security/news/2010/10/it-is-possible-to-kill-the-evercookie.ars

[15] S. Kamkar, "Evercookie - Never Forget," *New York Times*, 2010. [Online]. Available: http://www.nytimes.com/2010/10/11/business/media/11privacy.html

[16] P. Eckersley, "How unique is your web browser?" in *Privacy Enhancing Technologies*. Springer, 2010, pp. 1–18. [Online]. Available: http://www.springerlink.com/index/0J1M07443GU00H07.pdf

[17] H. Blodget, "Compete CEO: ISPs Sell Clickstreams For 5 Dollars A Month." [Online]. Available: http://seekingalpha.com/article/29449-compete-ceo-isps-sell-clickstreams-for-5-a-month

[18] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," in *Proceedings of the 13th conference on USENIX Security Symposium-Volume 13*. USENIX Association, 2004, pp. 21–21.

[19] Electronic Frontier Foundation, "Tor Project: Overview." [Online]. Available: https://www.torproject.org/about/overview.html.en

[20] M. Reed, P. Syverson, and D. Goldschlag, "Anonymous connections and onion routing," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 4, pp. 482–494, 1998.

[21] B. Levine, M. Reiter, C. Wang, and M. Wright, "Timing attacks in low-latency mix systems," in *Financial Cryptography*. Springer, 2004, pp. 251–265.

[22] D. Herrmann, R. Wendolsky, and H. Federrath, "Website fingerprinting: Attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier," *Proceedings of the 2009 ACM Workshop*, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?id=1655013

[23] S. Newton, "Ant Video Downloader Firefox Addon Tracking My Browsing," 2011. [Online]. Available: http://iwtf.net/2011/05/10/ant-video-downloader-firefox-addon-tracking-my-browsing/

[24] Microsoft, "Internet Explorer 9 Tracking Protection Lists." [Online]. Available: http://ie.microsoft.com/testdrive/Browser/TrackingProtectionLists/

[25] Google, "Google Public Policy Blog: Keep your opt-outs." [Online]. Available: http://googlepublicpolicy.blogspot.com/2011/01/keep-your-opt-outs.html

[26] IETF, "Do Not Track: A Universal Third-Party Web Tracking Opt Out." [Online]. Available: http://tools.ietf.org/html/draft-mayer-do-not-track-00

[27] ZDNet.de, "Briten verabschieden Regelung zum Schutz vor Web-Tracking." [Online]. Available: http://www.zdnet.de/news/digitale_wirtschaft_internet_ebusiness_briten_verabschieden_regelung_zum_schutz_vor_web_tracking_story-39002364-41551921-1.htm

[28] R. Steidle and U. Pordesch, "Im Netz von Google. Web-Tracking und Datenschutz," *DuD Datenschutz und Datensicherheit*, p. 324, 2008.

[29] J. Letzing, "Web tracking bill draws fire from Facebook, Google." [Online]. Available: http://www.marketwatch.com/story/web-tracking-bill-draws-fire-from-facebook-google-2011-05-03

[46] http://ec.europa.eu/justice/policies/privacy/workinggroup/index_en.htm

[47] Similar efforts have been started in the U.S.: http://wyden.senate.gov/newsroom/press/release/?id=0b8d693f-7cab-4ba4-aae6-42b66b1eef0e

[30] B. X. Chen, “iPhone tracks your every move, and theres a map for that,” 2011. [Online]. Available: http://www.wired.com/gadgetlab/2011/04/iphone-tracks

[31] M. Panzarino, “It’s not just the iPhone, Android stores your location data too,” 2011. [Online]. Available: http://thenextweb.com/google/2011/04/21/its-not-just-the-iphone-android-stores-your-location-data-too/

[32] C. Carmy, “Foursquare dreht auf,” *Technology Review*, 2011. [Online]. Available: http://www.heise.de/tr/artikel/Foursquare-dreht-auf-1249969.html

[33] ARTICLE 29 Data Protection Working Party, “Opinion 13/2011 on Geolocation services on smart mobile devices,” 2011. [Online]. Available: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp185_en.pdf