Privacy Awareness about Information Leakage: Who knows what about me?

Delfina Malandrino
ISISLab, University of Salerno
84084, Fisciano (SA), ITALY
delmal@dia.unisa.it

Andrea Petta
ISISLab, University of Salerno
84084, Fisciano (SA), ITALY
andrpet@gmail.com

Vittorio Scarano
ISISLab, University of Salerno
84084, Fisciano (SA), ITALY
vitsca@dia.unisa.it

Luigi Serra
ISISLab, University of Salerno
84084, Fisciano (SA), ITALY
luigser@gmail.com

Raffaele Spinelli
ISISLab, University of Salerno
84084, Fisciano (SA), ITALY
spinelli@dia.unisa.it

Balachander Krishnamurthy
AT&T Labs-Research
bala@research.att.com

ABSTRACT

The task of protecting users' privacy is made more difficult by their attitudes towards information disclosure without full awareness and the economics of the tracking and advertising industry. Even after numerous press reports and widespread disclosure of leakages on the Web and on popular Online Social Networks, many users appear not be fully aware of the fact that their information may be collected, aggregated and linked with ambient information for a variety of purposes. Past attempts at alleviating this problem have addressed individual aspects of the user's data collection. In this paper we move towards a comprehensive and efficient client-side tool that maximizes users' awareness of the extent of their information leakage. We show that such a customizable tool can help users to make informed decisions on controlling their privacy footprint.

Keywords

Privacy awareness, Information Leakage, Privacy-enhancing Technologies

1. INTRODUCTION

Given the increasingly important role of online communication in people's everyday life, enhancing users' privacy protection is a critical issue. Increasing amounts of both personally identifiable information (PII) and sensitive (e.g., medical, financial and family) information continue to be leaked [12, 14, 19]. The situation has been exacerbated through the introduction of free popular services, such as on Online Social Networks (OSN), and the ability of advertising companies to deliver targeted advertising. Privacy can be undermined by third parties [5]. Users effectively pay for these free services through micro payments of ever-greater amounts of personal information.

Different meanings and dimensions of privacy have been discussed in literature [6]. We will adopt the definition of privacy as the "*right to prevent the disclosure of personal information*" [30] that stems out from an 1890 definition of privacy as the "*right to be let alone*" [29]. In a taxonomy of privacy violations [26], four groups of activities have been recognized as harmful for both daily life and online privacy of individuals. They are "*Information Collection*" which includes all activities related to surveillance, "*Information Processing*" which involves the way information is stored, aggregated, linked and used, "*Information Dissemination*" which involves the way information is distributed and "*Invasion*", with intrusions into people's private affairs.

Overall, collection, processing and dissemination of personal information can raise serious privacy issues among users when they go online, for a variety of daily activities such as online banking, business transactions, online shopping, social network interactions and so on.

The online marketing methods of network advertisers, for example, have given rise to concerns about user's privacy [4]. Although the practice of tracking individuals' online activities increases the effectiveness and the revenues of the marketers' campaigns, it also undermines the privacy of users, mainly because it relies heavily on users' personal information. Pseudo-anonymous data collected and linked with PII such as email addresses and credit card number, may be sold by aggregators. The possessors of such data may use it for identity theft, social engineering attacks, online and physical stalking and so on [7, 10, 15, 23].

This paper makes several contributions. First, we discuss some of the most important requirements that tools have to exhibit to protect users' privacy on the Web, that is: comprehensiveness, support and awareness, performance and effectiveness. Second, we show how NoTrace [16, 17], a privacy-enhancing tool: (1) Fully addresses the aforementioned requirements (2) Displays in real time, that is during a browsing session, leakages of personal information (3) Raises awareness of measures to safeguard personal data and search habits (4) Improves privacy of Web users. Third, we show that NoTrace can detect *more* information leakage than other popular privacy tools at a *lower cost*. Fourth, we design a hierarchy of the most important privacy threats analyzing the ways in which personal and sensitive information

are sent to third party sites. We derive an ordering of the importance of the tools according to the countermeasures they provide and their effectiveness in limiting the disclosures of important information. Fifth, we show that, by linking pieces or bits of personal information leaked towards different third party sites, it is possible to identify users and derive their interests and browsing habits. We show how NoTrace is able to give real time information about which aggregators have what portion of users' personal data.

2. PRIVACY AWARENESS AND NOTRACE

We summarize privacy awareness as encompassing the perception of: (1) *Who* is tracking, receiving or collecting private information (2) *When* information is collected (3) *What* information other entities receive, store and use (4) *How* pieces of information are processed, linked and aggregated to potentially build detailed users' profiles.

Although the complexity and the efficacy of data mining technologies are growing quickly to increase the effectiveness of behavioral advertising, the awareness of privacy erosion is growing slowly [12]. In this paper we show how NoTrace informs users about which pieces of personal information are disclosed to third party entities. As more users learn about their information leakage they may be able to make better decisions about controlling their privacy [18].

Tools for privacy protection should exhibit important requirements as discussed by PÃűtzsch in [24]. We will briefly discuss some of them here, highlighting in the next Sections, how NoTrace is able to address them and the necessary changes we made to the tool since our earlier works [16, 17]. These important requirements include:

- **Offer support, no assumption of responsibility** helping users to make informed decisions.
- **Comprehensiveness** in terms of threats to address and corresponding countermeasures to provide.
- **Awareness and full control** over privacy leakage and countermeasures to adopt to limit its diffusion.
- **Performance and effectiveness** in order to make the tools longer used by users, since excessive delays involve an abandonment by users after first use [9].

NoTrace, a Mozilla Firefox add-on included in the Privacy & Security Category of the Mozilla Community[1][2], relies on a modular architecture. This modularity represents the key factor to provide measures for privacy protection for many privacy threats, by also guaranteeing efficiency and effectiveness. From the technical point of view, NoTrace leverages the Cross Platform Component Object Model (XPCOM) framework[3], that allows the development of modular software and provides tools to create, assemble and manipulate components at run-time. Specifically, we have implemented three different components: the first implements techniques that manages HTTP requests/responses headers; the second component manages the filtering mechanisms we developed to apply on-the-fly transformations before the browser rendering begins; finally, the third component manages the traditional URL-based blocking mechanism. The integration of new countermeasures can be realized by implementing the new technique as part of one of the three NoTrace components (chosen according to the kind of resources to manipulate) by only developing the JavaScript functions that represent the logic of the new functionality to offer to protect privacy. Otherwise, whether new types of mechanisms are required, it is possible to design and implement a new XPCOM component, that has to be registered into Mozilla, and has to implement the needed interfaces and the corresponding methods (i.e., nsIObserver and nsISupports).

[1] `https://addons.mozilla.org/en-us/firefox/addon/notrace/`

[2] Development version available at: `http://www.isislab.it/projects/NoTrace/Download/Dev/notrace@unisa.it.xpi`

[3] `https://developer.mozilla.org/en/XPCOM`

2.1 Privacy support and comprehensiveness

NoTrace supports users' needs through several privacy settings that can be fine-tuned according to experience and expertise [17]. It is able to address several online privacy threats by providing, all in one, opportune countermeasures, whereas each of them is singularly provided by other popular add-on in this field, that sometimes are in conflict with each other[4]. It also provides countermeasures to many privacy threats with no conflicts and performance slowdown of the browser, that may occur if multiple tools have to be installed to provide the same countermeasures. An experiment verifying this claim is discussed in Section 3.1.

We extended the set of provided measures described in our earlier works [16, 17] with techniques to block requests for large advertising companies, or to alter the browser fingerprint information [8] for requests to third party sites. Additionally, we implemented a new *"External-filtering"* mechanism, that is able to access to the stream of bytes received by the browser immediately before the rendering of the Web page. This new filtering mechanism has been used to implement new protection measures, such as those that look at Cookies and Referer fields set in external JavaScript codes. The combination of HTML inspection via the publish/subscribe design pattern (to implement the Content-based mechanism [17]) and external filtering via the nsITraceableChannel and nsIStreamListener interfaces (to implement the External-filtering mechanism), both available via the Mozilla API, allows NoTrace to access the HTTP stream before the browser. No other tool has harnessed this combination before, leveraging, therefore, URL-based filtering mechanisms, only.

2.2 Awareness and full control

To educate users about what personal and sensitive information they leak towards third party aggregators and the information that is inferred based upon their behavior, we deployed in NoTrace specific awareness modules. Specifically, NoTrace shows which information are leaked towards third party entities, for each visited Web site. It also allow users to be informed about which information they leak towards the most popular aggregators and advertising companies (see Fig. 1). Therefore, users may be informed about which fraction of their personal data is known and shared by many popular third party entities.

No other tool has envisioned and provided this type of awareness, about personal information leakage, to Web users.

[4] Ghostery message: "*Warning! When combined with other cookie monitoring addons such as Beef Taco, Cookie Monster, and Google Opt-Out, this feature can cause unresponsive script errors. If you experience this error, please try disabling this feature or conflicting addons*".

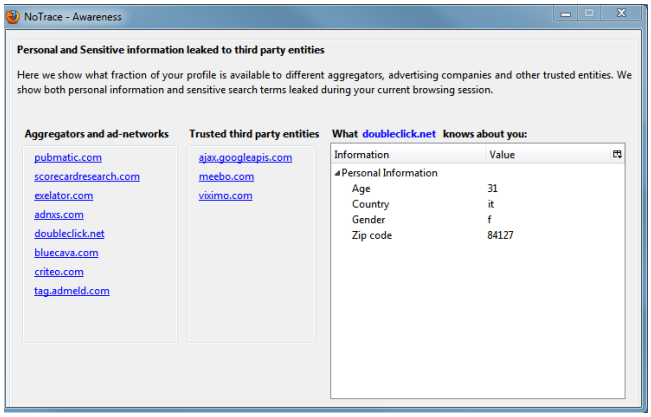

Figure 1: How NoTrace makes users aware of what third party entities know about them. In addition (not shown in the figure), full name and email address bits are leaked towards bluecava.com, while the Social Status (i.e., Single) is leaked towards exelator.com. Both entities are advertising companies.

2.3 Performance and effectiveness

Excessive delays experienced by users when using a tool may involve its abandonment just after first use [9]. We tested the effectiveness and the impact on the user's experience of this improved version of NoTrace, because of some changes we made in this work to speed up performance. The positive results are presented in Section 3.1.

3. A COMPARATIVE STUDY

We compare NoTrace and other popular tools that are comparable in terms of functionalities: AdBlock Plus[5], NoScript[6], Ghostery[7], and RequestPolicy[8]. A summary of their functionalities and main characteristics are shown in Tables 1 and 2, respectively. Table 1 shows that all tools provide functionalities to filter advertisements and to block third party requests. Among the analyzed techniques to protect privacy of individuals NoTrace fully support them while they are partially supported by the other tools. Table 2 highlights the differences between NoTrace and the other tools in terms of provided Awareness, Crowdsourcing filtering rules, Blocking methods and Configuration properties.

Our comparative study will cover both the impact on users' perceived experience and performance (Section 3.1) and the effectiveness of the tested tools in terms of false positives (FP) and false negatives (FN) due to the filtering rules (Section 3.2). We show that NoTrace provides privacy protection at a lower cost and without degrading page quality or cause functional breaks of the returned Web pages.

3.1 Impact on User Experience

Following [11], our data set consists of the top-100 Web sites from 15 Alexa categories[9]. As measuring methodology to gather realistic data about page downloads, we augmented the Firefox browser by the Pagestats extension[10]. This extension was used to retrieve 1500 pages and involved over 200,000 URLs to be analyzed. To allow for a fair comparison we configured the tools so that they provide the same functionalities. We enabled in NoTrace, for all experiments performed, the techniques that filters out advertisements, Web bugs, hidden third party scripts, and the technique that blocks requests for third party domain servers and aggregators. The result of the behavior of each tool is compared individually to the result of the experiment *without* any tool installed, named "NoAddons", that represents our baseline measurement. We performed five different tests, one per each tool appropriately configured. To automate tests and avoid interferences among them, we also used different browser profiles. We performed experiments sequentially, so that the influence of the turnover request is minimized. Tests were performed on a PC Intel(R) Core i7-2600@3.40 GHz with 8GB RAM and 64 bit Windows 7 Operating System.

3.1.1 Response time results

We compared how the tested tools perform in terms of mean response times when applying the filtering capabilities on our data set. We calculated the gain in terms of response time when third party objects are being removed from users' requests. We computed the objects retrieved on a page when filtering is applied, against objects retrieved under normal conditions (i.e., the "NoAddons" experiment).

NoTrace shows better behaviors than those exhibited by AdBlock and Ghostery, but it has a greater response time when compared with NoScript and RequestPolicy (overhead of almost 600ms for both). Specifically, NoTrace is able to save (on average) about 1.9 seconds against the baseline (3832ms vs. 1940ms). Additionally, it is able to block unwanted objects and save 35% of the total MegaByte transferred in downloading Web pages. The saved bytes for NoScript, RequestPolicy, AdBlock Plus and Ghostery are 54%, 57%, 23%, 29%, respectively.

The principal reason why NoScript and RequestPolicy is faster is the large number of resources blocked via their filtering rules. NoScript blocks, regardless of the real danger of detected objects, *all JavaScript code*, even those that are essential to the correct behavior of the page, while RequestPolicy has a stricter set of rules, avoiding the page break for very popular Web pages (e.g. YouTube) only because they are included by default in the startup whitelist. We show empirically in Section 3.2 that the NoScript and RequestPolicy strict policies negatively impact the quality and the functionality of the Web pages returned, drastically compromising the user's Web experience.

3.1.2 Browser performance results

Among the studied privacy protection tools, none is able to fully address all known privacy threats. A more privacy focussed navigation would require the installation and configuration of many of them into the browser, leading to possible performance degradation when a browser loads and interacts with multiple add-ons[11]. To study this we compared the performance of Firefox when loading up to 8 addons (i.e., AdBlock Plus, NoScript, Ghostery, RequestPolicy,

[5] http://adblockplus.org
[6] http://noscript.net
[7] http://www.ghostery.com
[8] https://www.requestpolicy.com
[9] http://www.alexa.com/
[10] http://www.cs.wpi.edu/~cew/pagestats/
[11] http://blog.mozilla.org/addons/2010/06/14/improve-extension-startup-performance/

Table 1: Summary of supported functionalities.

Tool	HTTP Header removal	3rd party cookies	Flash cookies	Web bugs	HTML5 Local Storage	Opt-out cookies	3rd party requests	Ads	3rd party script execution
NoTrace	√	√	√	√	√	√	√	√	√
Ghostery	–	–	√	√	–	√	√	√	√
AdBlock Plus	–	–	–	∼[a]	–	–	√	√	√
NoScript	–	–	–	∼[a]	–	–	√	√	√
RequestPolicy	–	–	–	∼[a]	–	–	√	√	√

[a]This threat may be blocked as consequence of the application of other functionalities

Table 2: Summary of the main properties of the tested privacy tools. Refer to Section 2 to recall the meaning of the different Blocking methods.

Tool	Awareness: Blocked URLs	Awareness: Data leakage	Crowdsourcing filtering rules	Blocking methods	Configuration properties
NoTrace	√	√	√	URL Content External	*Configuration*: Checkbox to activate/deactivate techniques *Extra Step:* Whitelist, Feedback by users, Crowdsourcing of rules
AdBlock Plus	√	–	√	URL	*Configuration:* Loading of subscription lists *Extra Step:* Adding on-the-fly new regular expressions to filter unblocked objects
Ghostery	√	–	–	URL	*Configuration:* Checkbox to activate or deactivate techniques, block/unblock *ad*-companies *Extra Step*: Selectively block a specific *ad*-company
NoScript	√	–	–	URL	*Configuration*: Checkbox to activate or deactivate techniques *Extra Step*: Whitelist, Blacklist and Embedding Objects to configure (temporarily or permanently)
RequestPolicy	√	–	–	URL	*Configuration*: Loading cross-site whitelists *Extra Step*: Add pairs of domains for which requests are allowed. Selectively enable/disable filtering on-the-fly for a Web site

Taco, RefControl, Privacy Choice and TrackMeNot) with specific techniques (i.e., advertisements and Web bugs filtering, third party JavaScript code execution blocking, opting-out from the tracking performed by *ad*-networks, HTTP Referer blocking) against its performance when only NoTrace is loaded as a way to provide "all in one" functionality.

Results of this experiment showed that, on average, Firefox loading time was 1260 milliseconds for the multiple-installation against 360 milliseconds for NoTrace, showing an evident gain in terms of startup time.

We also tested the memory footprint during a reasonable facsimile of several hours of Web browsing. We ran the MemBench script[12], which is a memory test benchmark that opens 150 popular Web sites, one per tab. We used the Mozilla Firefox "about:memory" monitoring tab to measure memory consumption at the startup, with the 150 tabs open and then after closing them. The metric we measured was *resident* memory consumption, which is the amount of physical memory used by Firefox, measured by the operating system. As a result, after closing 150 tabs, Firefox resident memory consumption with multiple extensions is 2.8x larger than Firefox with only NoTrace installed without any change to the add-ons. After closing those tabs, the initial memory allocated to Firefox has not fully released, and, instead, it increased up to double the initial value.

We also analyzed the memory consumption separately experienced by each tool. AdBlock Plus starts with the highest allocated memory since it needs to load in memory the subscription list. Ghostery shows worse performance since the resident memory at the end of the experiment was 4 times higher than the startup value. Finally, NoScript and RequestPolicy show better memory consumption values due to the high number of blocked resources.

We also tested how *multiple installations* of tools may involve a larger consumption of the memory. The final allocated memory for the *NoTrace single installation* was 120MB. Installing and using three tools, that is NoTrace, AdBlockPlus, and Ghostery, involved an increase, on average, of the value of the not released memory up to 300 MB. We have not included RequestPolicy and NoScript in this test, since the amount of the resident memory and of the final allocated memory drastically decreased, due to the number of resources that they block and not because of better performance.

Overall, by using NoTrace alone we can save on average 60% of the memory; an amount that becomes more significant in the mobile environment.

3.2 Effectiveness

This study was carried out by manually analyzing the top-10 Web sites from the Alexa News category[13], with nearly 1400 embedded resources. To differentiate between objects that are needed for the correct functioning of a Web page and objects that should be filtered out, we manually added all needed CDNs domains to the whitelist. We call this technique *intelligent filtering*.

3.2.1 Results

We analyzed False Positives (FP) and False Negatives (FN) for all tested tools. Due to space limitations we will

[12]`http://gregor-wagner.com/tmp/mem`

[13]Data retrieved on 20th October, 2012

discuss only NoTrace errors in detail. Table 3 shows in column 7, the number of FP detected when applying *intelligent filtering* (i.e., IF in Table 3) and without considering domains that serve their content for first party sites (i.e., NoIF). As an example, for the `foxnews.com` Web site, its content also comes from a third party entity, that is `fncstatic.com`, mostly serving Web images. Thus, by indiscriminately blocking all third party resources, the quality of the page could be degraded without any privacy improvement, leading to a large number of FP.

With *intelligent filtering* we will avoid all FP. The same argument applies to all the analyzed Web sites. NoTrace's FN, instead, can be due to: (i) First party requests for resources that are not available in the DOM (ii) Objects served by CDNs of first party sites, and (iii) 3d-party requests for resources that are not available in the DOM.

The first category includes requests for Web bugs, such as, the request for `us.bc.yahoo.com/b`, that is a 1x1 pixel GIF image hosted by `yahoo.com`. NoTrace is not able to block this, as its technique to filter out Web bugs looks at the height and weight properties of the Web images only available in the DOM of the requesting Web page. Similar to AdBlock Plus, we allow users to add an ad-hoc filtering rule to block it.

The second category includes errors due to the inclusion of the CDN servers into the whitelist because of their role in serving needed content for the requested Web pages[14]. Adding a CDN to the whitelist may result in allowing requests for unwanted third party objects. A modification to the whitelist's management is needed to inspect the potential harm of a third party object before checking the presence of the corresponding domain in the whitelist.

The third category includes errors due to requests directed to third party entities for resources not available in the DOM of the Web pages. Here, the high number of errors is due to a request for a JavaScript code[15] that loads a certain number of both harmless scripts and malicious scripts (13 out of 16 errors are Web bugs for the `weather.com` Web site). If we remove the loader we can avoid tracking, but at the same time, we may break the quality of the Web page, since the harmless scripts are used for page formatting and additional site's functionalities. A feasible solution requires inspecting the requested URL, extract the internal scripts, block the unwanted ones (wx-metrics.ts2.js, in the example above), and then resubmit the modified URL.

In summary, as shown in Table 3, the incidence of FP and FN for NoTrace is low, while as expected, NoScript and RequestPolicy exhibit the highest number of errors.

To compare tools, we also plotted the number of FP and FN of the analyzed sites. Fig. 2 shows NoTrace's better behavior and the worst behavior of both NoScript and RequesPolicy with an extremely high FP. RequestPolicy has over 100 errors, in two cases. Only in few cases NoScript and RequestPolicy exhibit a low FP as the corresponding domains have been whitelisted as a configuration default. Their high error rate is due to naively blocking all third party requests, leaving users to adjust the filtering, by whitelisting URLs, or disabling filtering on a specific site when the quality appears degraded. Properly configuring the whitelist requires substantially more expertise than an average user can reasonably be expected to have.

[14] `http://i.cdn.turner.com/cnn/.e/img/3.0/1px.gif`

[15] `http://d.imwx.com/jsRollup?rollup=/managedfe/js/TWC/util/\\social-loader.js,/managedfe/js/TWC/util/rotatingLogo.js/\\managedfe/js/TWC/util/social-share.js/...`

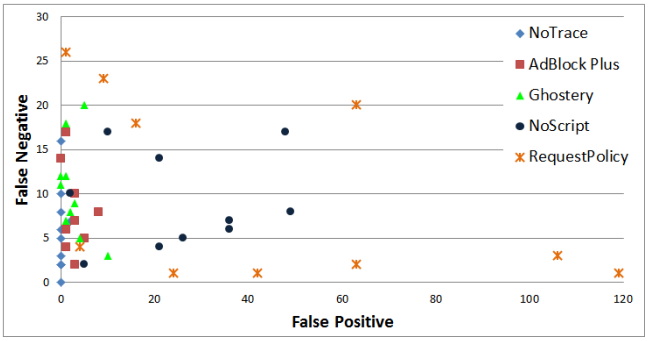

Figure 2: Comparison of tools in terms of FP and FN after blocking.

As further evidence of this claim, we examined what domains need to be whitelisted for various actions a user would typically do on a popular website: online.wsj.com. While NoTrace, Ghostery and AdBlock Plus did not degrade usability of the site, NoScript and RequestPolicy required multiple domains to be whitelisted (namely wsj.com, wsj.net, akamai.net, akamaihd.net etc.). Further details are shown in Appendix A.

4. INFORMATION LEAKAGE STUDY

We now explore the manners of leakage through which personally identifiable information and sensitive information are sent to *third party sites*, such as third party Cookies, Referer header, Web bug, third party JavaScript, Redirect Tracking, or advertisements. The "*Redirect Tracking*" leakage vector uses the HTTP redirect mechanism to redirect a user to the URL of a third party site. Based on the leaks we found, we classify the most popular threats for privacy and countermeasures to be provided by privacy preserving tools. Finally, we show that NoTrace is able to detect the most important leakages *and* at a lower cost.

4.1 Methodology

To analyze the leakage of personal and sensitive information we used 18 (sub)categories of Alexa and selected the top-10 sites that allow users to register. The categories are: Health, Travel, Employment, OSN, Arts, Relationships, News, PhotoShare, Sports, Shopping, Games, Computer, Home, Kids_and_teens, Recreation, Reference, Science, and Society. We extended the data set used in Section 3.1 to consider two categories–OSN and Relationships–with a large number of registered users, one–Employment–that involves users supplying private information, and one–PhotoShare–that may involve leaks due to potentially harmful specific actions, such as inputting content. We set up accounts with the corresponding first party sites rather than signing in via a third party account. We also enabled the option "Remember me" for sites that allowed that option, to study if private information are stored and then sent to third party sites.

We added detailed information to the 180 accounts we built, including full name, email address (required for all accounts), Date Of Birth (DOB), Social Security Number (SSN), zip code, home address, personal cellphone, school

Table 3: Effectiveness on popular Web sites: FP and FN. For NoTrace we also consider whitelisted CDNs.

Web Site	Web site's CDNs	AdBlock Plus		Ghostery		NoTrace		NoScript		RequestPolicy	
		FP	FN	FP	FN	FP (IF/NoIF)	FN	FP	FN	FP	FN
news.yahoo.com	yimg.com	1	10	10	3	0/10	0	10	17	9	23
edition.cnn.com	turner.com	0	34	1	12	0/3	2	21	14	1	26
weather.com	imwx.com	3	7	5	20	0/29	16	36	7	16	18
reddit.com	redditmedia.com redditstatic.com	3	3	2	8	0/2	3	5	2	24	1
my.yahoo.com	yimg.com yahoapis.com	3	5	3	9	0/2	7	2	10	4	4
bbc.co.uk/news	bbcimg.co.uk bbci.co.uk	1	8	0	11	0/14	5	36	6	106	3
foxnews.com	fncstatic.com	8	6	1	18	0/35	2	49	8	63	2
nytimes.com	nyt.com	1	11	0	12	0/7	10	48	17	63	20
huffingtonpost.com	huffpost.com	1	23	1	7	0/4	6	21	4	42	1
guardian.co.uk	guim.co.uk	5	10	4	5	0/3	8	26	5	119	1
Total		26	117	27	105	1/109	59	254	90	447	198
Recall/Precision		0.93/0.77		0.89/0.74		1.00/0.86		0.66/0.79		0.60/0.81	

and general education information, sexual orientation, political and intellectual beliefs, general interests (music, movies, and travel). They represent the bits of private information that may be leaked towards 3rd-party sites.

We then created a log of typical interactions between the user and the sites. We included actions that may uniquely identify the users from (a) search terms [2], (b) browser habits, (c) preferences about music, movie and books[16], and (d) the structure of their social networks [22]. We used the following six types of online users' interactions:

1. Account Login and Navigation. We logged in on all 180 sites and analyzed information leakage due to 3d-party cookies. We also visited 4 or 5 embedded links per page, to reflect typical navigation of a user [13].

2. Viewing/Editing Profile. To reflect the most common actions performed by users on OSN we analyzed the following actions: viewing one's own profile and editing it (`About` link in the profile page, "Write About Yourself"), viewing 5 friend's profiles, writing on the "Timeline" of 2 of them.

3. Searching the Web for Sensitive Terms. We examined seven sensitive categories of terms thought to be vital personal information that may be unwillingly disclosed while searching on the Web. We searched using `google.com` for terms in these seven categories: Health, Travel, Jobs, Race and Ethnicity, Religious beliefs, Philosophical and Political beliefs, Sexual orientation. For each search term we also navigated through the first 2 search result pages. The distribution of the 20 keywords across the 7 sensitive categories is the following: 3 for Health, 5 for Travel, 2 for Jobs, 2 for Race and Ethnicity, 3 for Religious beliefs, 4 for Philosophical and Political beliefs, and 1 for Sexual orientation.

4. Popular search. We chose 10 keywords from the top Google searches in 2012[17] and Google Trend Web pages[18].

5. Inputting content. Since leakage of private information can occur when users input content on Web sites, we analyzed the following actions: post and reply to questions on forums (2 actions), reply to dating messages (1 action), upload pictures (1 action).

6. Like-ing content. Leakage can occur through social plugins, such as Facebook Like Button, Google Plus Button, and so on. We analyzed the following actions: "Like" on Facebook (2 actions), "Share" via Facebook (2 actions), "+1" on Google Plus (2 actions), "Share" via Google Plus (2 actions). A recent study [27] showed that Facebook users exhibited a strong negative association between privacy concerns and engagement, i.e. posting, commenting and Liking of content.

We used Selenium[19] to automate tests, logging HTTP headers and saving both the HTML pages and JavaScript codes. We generated a set of strings that might be leaked to a 3d-party entity. The set included strings related to the personal information we added to the 180 accounts at their creation time, and the sensitive terms that we searched for. We searched the Selenium logs for these strings and removed false positives by hand. When leakage occurred, we recorded the leaked information, the manner of leakage, and the third party destinations.

4.2 Information leakage results

4.2.1 Categorization of the most important leakages.

By extending the work done in [12], we identified the following leaked bits (newly identified leakages are in bold): Full name, Email, ***IP address***, Country, Region, City, Zip code, ***Education*** and Employment, Gender, Age, ***DOB***, ***Interests*** (Movie and Music), ***Sexual orientation***, Political and religious beliefs and ***browser fingerprint*** information. Following the categorization discussed in that work, we organized these bits in three different categories: *High*, *Medium* and *Low*, taking into account their degrees of sensitivity and identifiability. In Table 4 we show the only *High* sensitive bit we discovered in our study, while Table 5 shows bits with *Medium* degree of sensitivity and identifiability. Results for the bits with *Low* sensitivity and identifiability are shown in Appendix B.

Given the vehicles through private bits can be leaked and the observed leakage, the tables show the count of the first party sites leaking the bits (column 2) and the number of 3d-party sites that receive the leaked bits (column 3).

A total of 44 first party sites out of 214 leaked private information. For the leakage in the seven categories of searches defined in Section 4.1, the number of the total first domains

[16] `http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf`

[17] `http://www.google.com/zeitgeist/2012/#the-world`

[18] `http://www.google.com/trends/hottrends`

[19] `http://seleniumhq.org/`

Table 4: Analysis of the most important manner of leakages for the *High Category*.

Types of sensitive searches	Leaking 1st party sites	Leaked to third party sites	Leakage Vehicles					
			Referer	Web bug	3d-party JS	Ads	3d-party cookie	Redirect Tracking
Health searches	3/4	24	75	8	5	3	0	0

Table 5: Analysis of the most important manner of leakages for the *Medium Category*.

Bits of private information	Leaking 1st party sites	Leaked to third party sites	Leakage Vehicles					
			Referer	Web bug	3d-party cookie	3d-party JS	Ads	Redirect Tracking
Email	10	7	3	24	1	1	0	0
Full Name	23	14	8	26	24	0	0	0
D.o.B	5	4	0	13	13	9	3	0
IP Address	41	61	34	48	80	42	20	0
Types of sensitive searches	**Leaking sites / Tot 1st party sites**	**Leaked to third party sites**	**Referer**	**Web bug**	**3d-party cookie**	**3d-party JS**	**Ads**	**Redirect Tracking**
Travel searches	6/7	13	18	16	0	5	0	0
Job searches	4/5	51	143	28	0	19	0	0
Religious Beliefs	3/5	12	41	1	0	1	0	0
Political Beliefs	2/4	12	33	4	0	7	0	0
Ethnicity	1/3	13	29	0	0	1	0	0
Sexual Life/orient.	1/3	1	0	0	0	1	0	0
Summary			309	160	118	86	23	0

involved is given by the Y value of the X/Y relation, shown in Tables 4 and 5.

Table 5 shows an important bit leaked by a number of sites to be the user's full name. This leakage raises concerns when this bit is combined with sensitive terms. Health terms are leaked in 3 of the 4 sites studied, as shown in Table 4. In Job and Travel searches, 4 out of 5, and 6 out of 7 of the studied sites show leakage respectively (actions discussed in Section 4.1). Health information could be combined with user's personal information and create difficulties while seeking health insurance. Job information combined with user's personal information can lead to privacy attacks such as identity theft[20].

Examining both tables together, we can see that the most important vehicles through which all types of categories' bits are leaked are the Referer HTTP field and Web bug; blocking them would yield better privacy protection.

In Fig. 3 we show the distribution of the most important privacy leakage vehicles across the Low, Medium, and High categories. We highlight the new leaked bits discovered in our study in bold as compared to [12]. For all categories the Referer is the most used vehicle to track users. Only for the Low category we saw differences across the 6 manner of leakages.

4.2.2 *Classification of the tools to improve privacy.*

Tables 4 and 5 are obtained by navigating without any privacy protection tool. We repeated the same automated interactions from Section 4.1 when privacy tools are used.

We found Ghostery still leaks full name, city, zip code, region, gender, age, DOB, IP Address and browser fingerprint. NoScript and RequestPolicy had less leakage, since their stricter filtering rules. Specifically, NoScript leaked zip code, gender, age, IP Address, while RequestPolicy full name, region and DOB.

Table 6 shows results of the effectiveness of the tools in limiting the disclosure of sensitive terms searched online. Due to space limitation, the "Header Leakage" column merges leakages via Cookie and Referer, while "URL Leakage" column merges the remaining 4 leakages.

NoTrace is most effective in reducing the diffusion of both personal and identifiable information and sensitive search terms, as the zero entries show.

[20] `http://www.job-hunt.org/privacy.shtml`

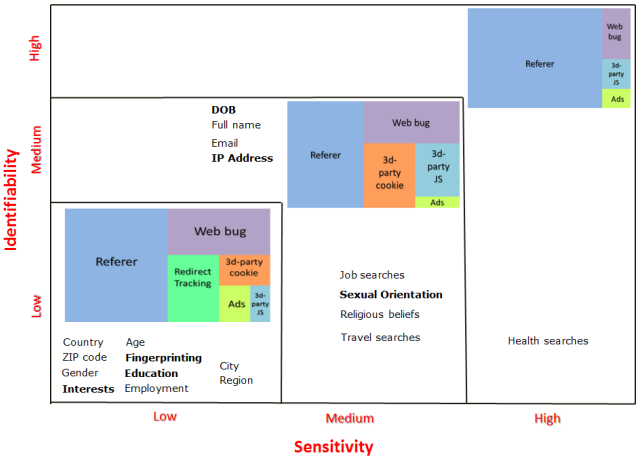

Figure 3: Distribution of the most important privacy leakage vehicles across the Low, Medium, and High category. The top-most right bits are highly sensitive and may quickly lead to identify users.

5. WHO KNOWS WHAT

We now see if it is possible to build a detailed profile about users by collecting and linking private information bits that users disclose online from diverse sources. Could the top aggregators find out precisely who a user is? To analyze what fraction of a user's profile is known by the top-10 aggregators, we instrumented Selenium to perform specific actions. Similar to the experiment described in Section 4, we carried out several actions that may uniquely identify users from

Table 6: Number of sensitive terms leaked when using a privacy protection tool

	NoTrace		AdBlock		RequestPolicy		NoScript		Ghostery	
	Header Leakage	URL Leakage	Header Leakage	URL Leakage	Header Leakage	URL Leakage	Header Leakage	URL Leakage	Header Leakage	URL Leakage
Health searches	0	0	0	0	2	1	11	4	14	7
Travel searches	0	0	7	1	3	0	7	0	8	8
Job searches	0	0	34	2	3	2	8	2	30	7
Religious Beliefs	0	0	10	1	0	0	2	0	6	3
Political Beliefs	0	0	23	6	0	0	2	0	34	22
Ethnicity	0	0	8	0	3	0	3	0	10	1
Sexual Life/Orient.	0	0	0	0	0	0	0	0	0	1
Summary	0	0	82	10	11	3	33	6	102	49

their interactions on the Web. They were (1) Logging into all 180 accounts (2) Viewing and editing all 10 profiles from the OSN category, post comments, post messages, share documents with "friends" (3) Search on all 10 shopping sites from the Shopping category, add items to shopping carts (without payment), create lists, "Like" content (4) Search on all 10 Job-related sites from the Employment category, sign up for email alerts (5) Search on all 10 Health sites from the Health category, post comments (6) Search on all 10 Travel site from the Travel category, book travel arrangements (without payment), visit Google maps site for itineraries, share with friends (via email and OSNs) (7) Reply to messages on five out of 10 Web sites of the Relationships category, that not required a Premium account (8) Create Photo Galleries on the `photobucket.com` Web site, upload images, add comments, share with friends, "Like" content (9) Watch videos on the `youtube.com` Web site, post comments, share with friends, "Like" content (10) Play songs on the `last.fm` Web site, post comments, share with friends, "Like" content.

All interactions were logged by Selenium during the experiment. We then inspected the Selenium logs to see if we could find any evidence of users' private information being leaked and which fractions of this information is known by the top-10 aggregator servers.

5.1 Results

We used the top-10 leak recipients identified in our data set (Table 7). To analyze results we used the same method of Section 4.1. We extended the set of strings to also look at sensitive Health terms (i.e., Pregnancy, Depression, Breast Cancer), Job terms (i.e., Analyst, Senior Analyst in New York), Travel terms (i.e., traveling from Napoli Capodichino to New York (JFK) and corresponding dates), music, book, and movie interests (i.e., Black Eyed Peas, Internet Traffic Measurement, and Viva l'Italia movie).

In Table 7 we report which bits are received by each aggregator. The fraction of known bits, i.e. number of received bits respect to the total number of analyzed bits, ranges between 12% for pubmatic.com, to 87% known by google-analytics.com. Surprisingly, the Health terms are leaked to almost all top-10 aggregators. Google Analytics is the top recipient of the leakages, since it receives 87% of leaked bits.

Linking of several different exchanges, *ad*-servers, or *ad*-networks (i.e., daisy chaining[21]) can increase chances of building detailed dossiers about users. We found in our study many communications among aggregators with leakage of private information. Column 2 of Table 8 shows the first party sites contacted. Columns 3 and 4 show the third party aggregators involved in daisy chaining. The last column shows the personal and sensitive information leaked in that process.

Daisy chaining is identified by examining the HTML body which includes an IFRAME that automatically triggers a request to the first aggregator. The aggregator's response includes a JavaScript file which triggers a request to the second aggregator. Linkage between the aggregators can be seen via the Referer header.

As last experiment, we saw if users' habits influence NoTrace's effectiveness in reducing the information leakage, by simulating 100 different random navigation behaviors. Each navigation behavior has a navigation part and a Web search part. For the navigation part, we chose at random a set S consisting of 2 to 5 Alexa categories. For each category, we selected five random Web sites to log in and visit, while we visit the remaining 5 sites without signing in.

For the Web search part we first defined lists of popular search terms[22], one for each Alexa category defined in Section 4.1: each list will contain terms to search on `google.com` *relevant* to that category. Then, we used Google to search three terms, selected uniformly at random from the lists of popular terms relevant for each of the categories in S, therefore from 6 to 15 terms. Further two terms to search are chosen uniformly at random from the remaining lists, i.e., for the categories not in S.

Results show that, regardless of the attitudes of the users while navigating the Web, the effectiveness of NoTrace is still high, as it is able to effectively prevent any information leakage. The results are not shown here because of zero values for both personal and sensitive information leakages.

6. RELATED WORK

Many technologies exist to protect the privacy of users when they are online. RequestPolicy and Ghostery were explicitly designed with the privacy implications of third party requests as their focus. NoScript was introduced as a security measure while AdBlock Plus was meant to block advertisements. Other examples of browser extensions are RefControl[23] for filtering out Referer headers and Milk, that automatically rewrites HTTP cookies to strictly bind them to the first-party domains from which they were set [28]. Proxies [3] have been proposed. However, all these tools miss some of the important requirements, described in Section 2. Regarding performance aspect, only a simple experiment of

[21] http://www.masternewmedia.org/online-advertising-management-ad-network-defaulting-and-daisy-chaining-for-ad-revenue-optimization/

[22] `http://www.google.com/trends/explore`

[23] `http://www.stardrifter.org/refcontrol/`

Table 7: Building a profile from pieces of private and sensitive information.

Aggregator	Email	IP Address	Country/Region/ City	Zip Code	Gender	Age	DOB	Interests	Health/ Job	Religious/ Political	Sex Orient.	Travel	Known bits [%]
doubleclick.net	–	√	√/√/√	√	√	√	√	√	√/√	√/–	–	√	81
google-analytics.com	√	–	√/√/√	√	√	–	√	√	√/√	√/√	√	√	87
scorecardresearch.com	√	–	√/–/√	√	√	√	–	√	√/√	√/–	–	√	69
adnx.com	–	–	–/√/√	√	√	√	–	–	–/–	√/–	–	–	37
yieldmanager.com	–	–	–/–/√	√	√	√	–	–	√/√	√/–	–	–	44
2o7.net	–	√	√/–/√	–	√	–	–	–	√/√	–/–	–	–	37
crwdcntrl.net	–	–	–/–/√	–	√	√	–	–	√/–	–/–	–	–	25
pubmatic.com	–	√	–/–/–	–	–	–	–	–	√/–	–/–	–	–	12
2mdn.net	–	√	√/√ /√	√	√	√	–	–	√/–	–/–	–	√	56
imrworldwide.com	–	–	√/√/√	–	√	–	–	–	√/–	√/–	–	–	37

Table 8: Leakage of private information through daisy chaining.

Count	First party sites	Daisy Chaining		Bits leaked
		First Aggregator	Second Aggregator (Family)	
1	www.bebo.com	bluecava.com	addvisor.net (Targus Info)	Full name, Zip code
1	www.bebo.com	bluecava.com	e.nexac.com (Datalogix)	Full name, Zip code
2	barnesandnoble.com	doubleclick.net	2mdn.net (Google)	Gender
1	gamespot.com	doubleclick.net	2mdn.net (Google)	Gender
2	youtube.com	doubleclick.net	googlesyndication.com (Google)	Gender
3	www.datehookup.com	doubleclick.net	pubmatic.com (Pubmatic)	IP Address
2	www.datehookup.com	doubleclick.net	criteo.net (Criteo)	IP Address
1	it.bab.la	adv.adsbwm.com	bid.openx.net (openX)	Ethnicity
1	travelocity.com	doubleclick.net	yieldmanager.com (Yahoo!)	Travel date and itineraries
1	www.espncricinfo.com	doubleclick.net	2mdn.net (Google)	City
1	www.youtube.com	doubleclick.net	2mdn.net (Google)	Age, Gender
1	www.linkedin.com	doubleclick.net	2mdn.net (Google)	Zip code, Gender

correctness and impact on Web sites has been presented for RequestPolicy [25]. Studies focused on the third party Web tracking leverage surface crawling, i.e. visiting the home page of the sites without following other links.

In a multi-year study of 1,200 Web sites authors found increasing collection of information about users from an increasingly concentrated group of tracking companies [14]. Some studies focus on specific threats to privacy and the extent of their tracking practice. Flash cookies [21] were found on 20 of top-100 sites (with surface crawling). In a follow-up study [1], authors showed further evidence of the usage of Flash cookies and found sites that had HTML5 local storage and HTTP cookies with matching values. Another study [12] showed that fully 56% of the sites in their sample (i.e., top-100 Web site from 12 Alexa categories and sub-categories, with a surface crawl) directly leaked pieces of private information to 3d-party aggregators.

Defining and quantifying vectors for tracking consumers on the Internet was done recently by the UC Berkeley Center for Law and Technology[24]. This study found that the most popular 100 Web sites dropped thousands of cookies, and that 84.7% of them were third-party cookies. They also showed that the Flash cookies use is declining among the most popular Web sites while HTML5 LocalStorage is rising across the Web sites they analyzed. When compared with our work, this survey presents some limitations. The first is related to the employed methodology. Authors of this census, studied the top-100, top-1000 and top-25000 popular Quantcast Web sites with a crawl that only looked at up to 6 links being followed. Moreover, their crawling method did not considered any human action (e.g. adding items to a shopping cart) and did not follow links set by JavaScript code. Additionally, the crawler did not login and maintain an identity while traversing sites. A key difference of our work is the crawling methodology which followed many links, considered human actions, allowed login and maintained the identity while navigating the sites. This crawling methodology led to an entirely different analysis.

Another study [20] presents results about the effectiveness of 11 blocking tools at mitigating third party Web tracking. Three consecutive crawls of the Alexa U.S. top-500 were performed using the FourthParty Web measurement platform[25] to study the tools' effectiveness. The study showed that the most effective tool was a combination of community-maintained blocklists.

Beyond the surface crawling of these works we also considered the user's behavior as they navigate the Web.

Separately, we analyzed the leakage of both private and sensitive information, simulating the typical interactions of the online users. We showed the effectiveness, in terms of limiting the disclosure of private information, of several popular privacy tools, by deriving an ordering of the tools that better improve privacy of the individuals. Additionally, we examined what fraction of a user's profile is available to the different aggregators. We also studied leakages that may occur via communications among them. Finally, we performed all these studies leveraging NoTrace, a privacy protection tool, that is able to efficiently and effectively provide many measures to protect privacy online.

[24] http://www.law.berkeley.edu/privacycensus.htm

[25] http://fourthparty.info/

7. CONCLUSION

We show key characteristics essential for privacy protection tools: support for users and comprehensiveness, awareness and full control, and high performance. We showed that our tool, NoTrace, achieves these requirements. It provides users with the ability to monitor who has access to which personal information over time, what information can be accessed by 3d-party entities, and which bits of private information can be linked to one other to potentially trace back to the real identity of a user. By making tangible what happens behind the scene NoTrace empowers users with a clear overview of the availability of their personal information. Awareness has the potential to alert users to the corresponding privacy risks and help them in making informed decisions about feasible countermeasures.

Beyond enabling awareness, we showed that NoTrace provides several measures to limit the diffusion of both personal and sensitive information, with higher efficacy and efficiency as compared to its most popular competitors. We also explored the most popular vectors for tracking, and how NoTrace is able to display these activities to users, and limit the diffusion of their private information. Moreover, we showed that by reverse engineering what leakage is going to the top-10 aggregators, it is possible to discover what fraction of a user's profile is available to them. Our results show that one of the top-10 aggregator is able to collect 87% of a user's private information. Finally, unlike earlier work, we employed a crawling methodology that reflects users' real behaviors during online activities.

Ongoing works include the analysis of new privacy leakage vectors, such as risks in mobile environments and external applications, via a mobile version of NoTrace. We are working on evaluating whether privacy protection can also help reduce load and thus save, *en passant*, the energy needed to download and render Web pages on mobile devices. We performed preliminary tests to compare battery consumption with and without privacy protection, with encouraging initial results.

8. REFERENCES

[1] M. Ayenson, D. J. Wambach, A. Soltani, N. Good, and C. J. Hoofnagle. Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning. Technical report, University of California, Berkeley, 2011. http://ssrn.com/abstract=1898390.

[2] T. Barbaro, Michael; Zeller Jr. A Face Is Exposed for AOL Searcher No. 4417749, 2006. The New York Times.

[3] C. Canali, M. Colajanni, D. Malandrino, V. Scarano, and R. Spinelli. A Novel Intermediary Framework for Dynamic Edge Service Composition. *Journal of Computer Science and Technology*, 27:281–297, 2012.

[4] C. Castelluccia, M.-A. Kaafar, and M.-D. Tran. Betrayed by Your Ads! In *Proc. of the 12th Int. Conf. on Privacy Enhancing Technologies*, volume 7384 of *PETS'12*, pages 1–17. 2012.

[5] G. Conti. *Googling Security: How Much Does Google Know About You?* Addison-Wesley, 2008.

[6] DeCew Judith Wagner. *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology*. Cornell University Press, 1997.

[7] C. Dwyer. Privacy in the Age of Google and Facebook. *IEEE Technology and Society Magazine*, 30(3):58–63, 2011.

[8] P. Eckersley. How Unique Is Your Web Browser? In *Proc. of the 10th Int. Conf. on Privacy Enhancing Technologies*, PETS'10, pages 1–18, 2010.

[9] D. F. Galletta, R. M. Henry, S. McCoy, and P. Polak. Web Site Delays: How Tolerant are Users? *Journal of the Association for Information Systems*, 5(1), 2004.

[10] R. Gross and A. Acquisti. Information Revelation and Privacy in Online Social Networks. In *Proc. of the 2005 ACM Workshop on Privacy in the Electronic Society*, pages 71–80, 2005.

[11] B. Krishnamurthy, D. Malandrino, and C. E. Wills. Measuring privacy loss and the impact of privacy protection in web browsing. In *Proc. of the 3rd Symposium on Usable Privacy and Security*, SOUPS '07, pages 52–63, 2007.

[12] B. Krishnamurthy, K. Naryshkin, and C. E. Wills. Privacy leakage vs. protection measures: the growing disconnect. In *Web 2.0 Security and Privacy Workshop*, 2011. `http://www2.research.att.com/~bala/papers/w2sp11.pdf`.

[13] B. Krishnamurthy and J. Rexford. *Web protocols and practice: HTTP/1.1, Networking protocols, caching, and traffic measurement*. Addison-Wesley, 2001.

[14] B. Krishnamurthy and C. Wills. Privacy diffusion on the web: a longitudinal perspective. In *Proc. of the 18th Int. Conf. on World Wide Web*, pages 541–550, 2009.

[15] V. Lawton. Privacy Commissioner of Canada. Popular websites in Canada disclosing personal information. `http://www.priv.gc.ca/media/nr-c/2012/nr-c_120925_e.asp`, 2012.

[16] D. Malandrino and V. Scarano. Supportive, Comprehensive and Improved Privacy Protection for Web Browsing. In *PASSAT 2011*, pages 1173–1176, 2011. `http://www.dia.unisa.it/professori/delmal/papers/PASSAT11.pdf`.

[17] D. Malandrino and V. Scarano. Privacy leakage on the Web: Diffusion and countermeasures. *Computer Networks*, 57(14):2833 – 2855, 2013.

[18] D. Malandrino, V. Scarano, and R. Spinelli. How increased awareness can impact attitudes and behaviors toward online privacy protection. In *PASSAT*, 2013.

[19] H. Mao, X. Shuai, and A. Kapadia. Loose Tweets: An Analysis of Privacy Leaks on Twitter. In *Proc. of the 10th annual ACM workshop on Privacy in the electronic society*, WPES '11, pages 1–12, 2011.

[20] J. R. Mayer and J. C. Mitchell. Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, SP '12, pages 413–427, 2012.

[21] A. M. McDonald and L. F. Cranor. A Survey of the Use of Adobe Flash Local Shared Objects to Respawn HTTP Cookies. Technical report, CyLab, CMUs, 2011.

[22] A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *Proceedings of the 2009 30th IEEE Symposium on Security and Privacy*, SP '09, pages 173–187, 2009.

[23] D. Perito, C. Castelluccia, M. Kaafar, and P. Manils. How Unique and Traceable Are Usernames? In *In Proc. of the 11th Int. Conf. on Privacy Enhancing Technologies*, volume 6794, pages 1–17. 2011.

[24] S. PÃűtzsch. Privacy Awareness: A Means to Solve the Privacy Paradox? In *The Future of Identity in the Information Society*, volume 298 of *IFIP AICT*, pages 226–236. 2009.

[25] J. Samuel and B. Zhang. RequestPolicy: Increasing Web Browsing Privacy through Control of Cross-Site Requests. In *Proc. of the 9^{th} international Conference on Privacy Enhancing Technologies*, PETS '09, pages 128–142, 2009.

[26] D. J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review. GWU Law School Public Law Research Paper No. 129*, 154(3):477–553, 2006.

[27] J. Staddon, D. Huffaker, L. Brown, and A. Sedley. Are privacy concerns a turn-off?: Engagement and privacy in social networks. In *SOUPS*, pages 1–13, 2012.

[28] R. J. Walls, S. S. Clark, and B. N. Levine. Functional Privacy or Why Cookies are Better with Milk. In *Proc. USENIX Workshop on Hot Topics in Security*, 2012.

[29] S. D. Warren and L. D. Brandeis. The right to privacy. *Harward Law Review*, 4(5):193–220, December 1890.

[30] A. Westin. *Privacy and Freedom*. New Jork Atheneum, New York, 1967.

APPENDIX

A. EFFECTIVENESS STUDY

Table 9: Analysis of which of the embedded Web sites that are blocked by tools are needed for the functioning of the online.wsj.com Web site. Other tools are not shown since they do not impact on the usability of the Web site.

Tool	Blocked domains	Actions	Break	Needed domains
NoScript	Doubleclick, BlueKai, atdmt.com	Page formatting	√[a]	wsj.com
	ChartBeat, akamai.net, llnwd.net	Sign in	√	wsj.net
	Peer39, OutBrain, msn.com	Search	√	wsj.net
	Facebook, Google,Twitter	Browse a JS Menu	√	wsj.net
	wsj.com, wsj.net	Watch a video	√	akamai.net
		Share: Email/Twitter/FB	√	Facebook, Twitter
		View Market Data	√	wsj.net
		Social Widgets Panel	√	wsj.net
		Find Answers	√	wsj.net
RequestPolicy	Doubleclick, BlueKai, atdmt.com	Page formatting	√[b]	wsj.com
	ChartBeat, akamai.net, llnwd.net	Sign in	√	wsj.net
	Peer39, OutBrain, msn.com	Search	√	wsj.net
	Facebook, Google,Twitter	Browse a JS Menu	√	wsj.net
	quantserve.com, rubiconproject, imrworldwide.com	Watch a video	√	akamaihd.net
	dowjones.com, akamaihd.net, googlesyndication.com	Share: Email/Twitter/FB	√	Facebook, Twitter
	scorcardresearch.com, krxd.net	View Market Data	√	wsj.net
	wsj.com, wsj.net	Social Widgets Panel	√	wsj.net
		Find Answers	√	wsj.net

[a]Break of JavaScript and Stylesheet Dependency
[b]Inaccessible Web site

Table 9 shows the results of the experiment meant to show that both NoScript and RequestPolicy have a stricter set of rules for filtering. We visited the `online.wsj.com` Web site and we performed the common set of actions on that site, that is Sign in, Search, Watch a Video and so on (see the "Actions" column). We started with a blank whitelist and when performing the actions, we reported that some breaks occur ("Break" column) and which domains have to be whitelisted ("Needed Domains"), to avoid these breaks and re-establish the correct behavior of the site. While NoTrace, Ghostery and AdBlock Plus did not degrade usability of the site, NoScript and RequestPolicy required multiple domains to be whitelisted (namely wsj.com, wsj.net, akamai.net, akamaihd.net etc.).

B. INFORMATION LEAKAGE STUDY

Table 10: Analysis of the most important manner of leakages for the *Low Category*. The "Interests" row include leakage of preferences about movie and music interests.

Bits of private information	Leaking 1st party sites	Leaked to third party sites	Leakage Vehicles					
			Referer	Web bug	Redirect Tracking	3d-party cookie	Ads	3d-party JS
City	35	69	380	80	0	4	12	9
Employment	1	2	17	0	0	0	0	0
Education	1	2	17	0	0	0	0	0
Gender	20	39	103	39	7	0	37	25
Age	6	21	107	16	5	0	12	14
ZIP Code	37	49	128	16	0	113	8	4
Country	2	2	0	3	0	0	1	0
Region	9	13	26	8	0	0	1	0
Interests	3	5	21	0	0	0	1	2
Fingerprint	21	18	0	168	211	0	0	0
Summary			799	330	223	117	72	54

Table 10 shows that the most leaked bits include zip code, gender and city, that are useful for ad-companies to provide targeted advertising and for identity theft. Additionally, the most important manner of leakage is through the Referer Leakage vehicle.