Third-Party Web Tracking: Policy and Technology

Jonathan R. Mayer and John C. Mitchell
Stanford University Stanford, CA
{*jmayer,mitchell*}*@cs.stanford.edu*

***Abstract*—In the early days of the web, content was designed and hosted by a single person, group, or organization. No longer. Webpages are increasingly composed of content from myriad unrelated "third-party" websites in the business of advertising, analytics, social networking, and more. Third-party services have tremendous value: they support free content and facilitate web innovation. But third-party services come at a privacy cost: researchers, civil society organizations, and policymakers have increasingly called attention to how third parties can track a user's browsing activities across websites.**

This paper surveys the current policy debate surrounding third-party web tracking and explains the relevant technology. It also presents the *FourthParty* web measurement platform and studies we have conducted with it. Our aim is to inform researchers with essential background and tools for contributing to public understanding and policy debates about web tracking.

***Keywords*-Web privacy; third-party tracking; Do Not Track.**

I. Introduction

The web has evolved to facilitate development and delivery of webpages composed of content from multiple websites. HTML, JavaScript, and CSS impose no restrictions on a webpage including elements from, or even delegating complete control to, a wholly unrelated website.[1] These design choices have contributed to a host of well-known and well-studied security vulnerabilities, including cross-site scripting (XSS) [2] and cross-site request forgery (CSRF or XSRF) [3], [4], that enable an unauthorized and unrelated "third-party" website to retrieve information from or perform actions on the "first-party" website that the user has voluntarily interacted with.

This paper examines the privacy implications of the opposite case—where a first-party website authorizes a third-party website to learn about its users (Figure 1).[2] Third-party services bring tremendous value to the web: they enable first-party websites to trivially implement advertising, analytics, social network integration, and more. But they also give rise to privacy concerns: over the past several years, researchers, civil society organizations, and policymakers have called attention to the increasing trend of third-party websites recording and analyzing users' browsing activities across unrelated first-party websites ("third-party web tracking" or "tracking" for short).[3]

This paper is intended to comprehensively familiarize computer security and privacy researchers with current policy and technology research on third-party web tracking. Much of the discussion is based on recent results from a new dynamic web measurement platform, FourthParty. We begin by presenting by FourthParty. The remainder of the paper is organized into two parts on third-party web tracking: one on policy, and one on technology.

The policy part opens by reviewing why third-party web tracking gives rise to privacy concerns and ways in which policy might be structured to address those concerns. It then provides an overview of regulation and self-regulation in the U.S. and EU, explaining the current governmental and business approaches to mitigating privacy issues in third-party web tracking. The discussion next turns to tracking business models and trends. A final section frames the economic debate on third-party web tracking and notes gaps in the current literature.

The technology part starts by surveying stateful and stateless technologies that can be used to correlate users' activities across websites. It next provides an overview of technologies that enable the delivery of third-party services with lessened privacy risk. Last, it reviews the user choice and self-help technologies presently available, including opt-out cookies, blocking, and Do Not Track.

This paper has a secondary aim. Debates on how to respond to third-party web tracking are now occurring every day in Washington and Brussels. We hope that by systematizing knowledge on third-party web tracking for the computer security and privacy community, we will ensure that it is best able to assist policymakers in developing solutions that adequately balance privacy, commerce, and a thriving web.

II. FourthParty

A. Why Web Measurement?

We have found several advantages to placing web measurement at the center of our methodology.

[1] Content Security Policy [1] allows a website to opt into such restrictions.

[2] This paper focuses exclusively on the web. Third-party tracking is rapidly growing in the mobile application space [5], [6], where it likewise merits research attention.

[3] There is, at present, significant debate about the precise contours of a "third party" and "tracking." This paper focuses on the least ambiguous case: an unaffiliated website that collects a user's browsing history.

Figure 1. Third-party advertising, social, and video content on the New York Times website. Analytics content is not visible.

Web measurement provides objective, reliable evidence that both furthers public understanding and establishes a sound basis for policymaking.

Second, web measurement is fast. Many claims about specific tracking practices can be supported or rebutted with mere hours of web measurement work.[4]

Web measurement facilitates longitudinal study. Often the very same hardware and software can be reused to collect and analyze data even years apart.

Last, web measurement can often be automated. Once a generic measurement tool has been built, it can be trivially applied to millions of websites.

B. Design Principles

Prior work on third-party web tracking has largely taken one of three approaches to measurement: monitor network traffic (e.g. [10], [11], [12], [13]), manually inspect browser state (e.g. [8], [14]), or develop a custom tool for a specific measurement task (e.g. [15], [16], [17]).

We developed FourthParty around three design principles that improve on these approaches.

[4]For example, when an advertising network contested our discovery that it was "history sniffing" [7], we were able to secure independent confirmation from two other research groups the same day. When Ayenson et al. [8], [9] contacted us on a weekday afternoon about a web analytics company using multiple "supercookie" technologies (see Section VII-A), we were able to verify their findings by evening.

1) General-purpose instrumentation: By implementing comprehensive instrumentation and logging only once, FourthParty avoids the need for many purpose-built tools, decreases duplication of effort, and trims development time.

2) Production web browser: Building on a production browser allows reuse of existing add-ons, including for automation, and closely emulates real-world browsing.

3) Standardized log format: A standardized, easy-to-manipulate log format facilitates data sharing and cuts back on redundant data gathering.

C. Implementation

We implemented FourthParty as an extension to Mozilla Firefox. It currently instruments the browser APIs for HTTP traffic, DOM windows, cookies, and resource loads. FourthParty also instruments JavaScript API calls on the `window`, `navigator`, and `screen` objects using getters, setters, and ECMAScript proxies [18]. All events are logged to a SQLite database.

On many pages FourthParty does not perceivably slow down Firefox; on highly dynamic pages, it can increase page load time by roughly 2-3x. We plan to make substantial performance improvements in an upcoming revision.

D. Analysis with FourthParty

Analyzing FourthParty data is fast. All of the FourthParty results presented in this paper were generated with Python

scripts, most of which took seconds to execute on a consumer notebook with databases including visits to thousands of popular websites.

Analyzing FourthParty data is also easy for researchers who are already familiar with SQL syntax. For example, a query that counts Google Analytics reports with an `anonymizeIp` instruction (see Section VIII-B) is just:

```
select count(*) from http_requests
where url like
'%google-analytics.com/__utm.gif%&aip=1%'
```

See `http://fourthparty.info` for the FourthParty source and related resources.

We used FourthParty to conduct many of studies we present in this paper, including on information leakage (Section III-B3), tracking technologies (Section VII), and blocking tool effectiveness (Section IX-B).

Third-Party Web Tracking Policy

III. Privacy Problems

This section reviews the privacy implications of third-party web tracking and notes the range of policy responses. The discussion proceeds in four phases. First, it details the browsing history information that is available to third parties and how that information is identifiable. Second, it explains how third-party web tracking could harm users. Third, it reviews survey results consistently showing that users would prefer to not be tracked. Last, it details the policy positions that various stakeholders have adopted in response.

A. Information Available

Web browsing history is inextricably linked to personal information. The pages a user visits can reveal her location, interests, purchases, employment status, sexual orientation, financial challenges, medical conditions, and more. Examining individual page loads is often adequate to draw many conclusions about a user; analyzing patterns of activity allows yet more inferences.

When a first-party page embeds third-party content, the third-party website is ordinarily made aware of the URL of the first-party page through an HTTP referrer or equivalent.[5] If the page embeds a `script` tag from a third party, the third party will also often learn the web page's title from `document.title`. Some first parties will voluntarily transmit even more information.

Collection of sensitive personal information is not a hypothetical concern. In mid-2011 we discovered that an advertising network, Epic Marketplace, had publicly exposed its interest segment data, offering a rare glimpse of what third-party trackers seek to learn about users [7]. User segments included menopause, getting pregnant, repairing bad credit, and debt relief. Several months later we found that the free online dating website OkCupid was sending to the data provider Lotame how often a user drinks, smokes, and does drugs [19]. When Krishnamurthy et al. [10] tested search queries on ten popular health websites, they found a third party learned of the user's query on nine of them.

B. Identifiability

A web browsing history is often personally identified or identifiable. Narayanan [20] recently proposed a taxonomy of five ways in which a pseudonymous[6] browsing history might become identified. Note that pseudonymity is quite fragile in protecting identity: discovering a user's identity *once* in a pseudonymous system is sufficient to also identify past and future interactions with the user.

1) The third party is also a first party: The third party may be a first party in another context, where the user voluntarily provided her identity. Facebook, for example, has over 800 million users and enforces a requirement that users provide their real name to the service. When a page includes a third-party Facebook social widget, Facebook identifies the user to personalize the widget.

2) A first party sells the user's identity: Some first-party websites voluntarily provide ("leak") a user's identity to third parties for pay. Some have even made a business model of it, usually appearing as a free sweepstakes or quiz. Several advertising data providers (e.g. Datalogix [21]) buy identifying information, retrieve the user's dossier from an offline consumer database, and use it to target advertising.

3) A first party unintentionally provides identity: If a website puts identifying information in a URL or page title, it may unintentionally leak the information to third parties. In a 2011 paper [10], Krishnamurthy et al. examined signup and interaction with 120 popular sites for information leakage to third parties. They reported that an aggregate of 48% leaked a user identifier[7] in a Request-URI or referrer.

Using a similar methodology, we examined identifying information leakage on the Quantcast U.S. top 250 websites [19]. We were able to test signup and interaction with 185 of the sites; we found that a username or user ID was sent to a domain with a different public suffix + 1 (PS+1)[8] on 113 (61%) of the websites in our sample. The five most frequent recipients and most prolific senders of username and user ID are presented in Table I and Table II respectively.

In the majority of instances the username or user ID was part of a user profile URL or page title. A better practice

[5] Some third-party content reports a first-party page's URL as a parameter in a request.

[6] We consider a web tracking system to be pseudonymous if it allows, with moderate probability, correlation of web activities by a device or user.

[7] While there is room for further confirmatory research, there appears to already be substantial evidence that usernames and user IDs can trivially be used to identify a user [22], [23], [24], [25]. Some companies have already deployed username-based matching in their products, including for social user matching APIs (e.g. InfoChimps), creating user profiles (e.g. Spokeo), and recommending account linkage (e.g. Google Social Search).

[8] Public suffix + 1 is an alternative to top-level domain + 1 that is more accurate for purposes of privacy measurement. See [26].

Table I
THIRD PARTIES RECEIVING USERNAME AND ID ON 185 POPULAR SITES.

Third-Party PS+1	Websites Leaking Username or ID
scorecardresearch.com	81 (44%)
google-analytics.com	78 (42%)
quantserve.com	63 (34%)
doubleclick.net	62 (34%)
facebook.com	45 (24%)

Table II
POPULAR WEBSITES LEAKING USERNAME AND ID.

First-Party PS+1	Third-Party PS+1s Receiving Username or ID
rottentomatoes.com	83
cafemom.com	59
lyricsmode.com	54
ivillage.com	53
livejournal.com	53

would be to use a single URL for all users viewing their own profile, e.g. `http://example.com/self/`, and to never include the username or user ID in the page title. Several of the sites we contacted were willing to implement these fixes, but many more preferred the functionality, convenience, and aesthetic of a username or user ID in URLs. It seems quite likely the practice will persist indefinitely among even the most popular sites.

We also observed other forms of identifying information leak. For example:

- Viewing a local ad on the Home Depot website sent the user's first name and email address to 13 companies.
- Entering the wrong password on the Wall Street Journal website sent the user's email address to 7 companies.
- Changing user settings on the video-sharing site Metacafe sent first name, last name, birthday, email address, physical address, and phone numbers to 2 companies.

In all of these cases the identifying information was included as a parameter in a first-party URL. The better practice is to send identifying information as part of a POST request body so it will not inadvertently leak to third parties.

4) The third party uses a security exploit: A third party may exploit a cross-site security vulnerability on a first-party website to learn the user's identity. Narayanan has shown how inadequate frame busting can facilitate identifying a user [27]. Huang and Jackson more recently demonstrated practical user identification through Facebook and Twitter sharing widget clickjacking [28].

5) Re-identification: The third party could match pseudonymous browsing histories against identified datasets to re-identify them, much like Narayanan and Shmatikov did with the Netflix Prize dataset [29] and the Flickr and Twitter social graphs [30], and Acquisti et al. did more recently with personal photos on a dating site [31]. A third party might, for example, compare browsing activity to the times and locations of links publicly shared by Twitter users.

C. Possible Harms

The risk of harm to consumers from web tracking arises from myriad potential scenarios. Each *particular* scenario may have a low probability of occurring. But the chance of *some* scenarios occurring is substantial, especially when considered over time and across many companies.

When considering harmful web tracking scenarios, we find it helpful to focus on four variables. First, an *actor* that causes harm to a consumer. The actor might, for example, be an authorized employee, malicious employee, competitor, acquirer, hacker, or government agency. Second, a *means of access* that enables the actor to use tracking data. The data might be voluntarily transferred, sold, stolen, misplaced, or accidentally distributed. Third, an *action* that harms the consumer. The action could be, for example, publication, a less favorable offer, denial of a benefit, or termination of employment. Last, a particular *harm* that is inflicted. The harm might be physical, psychological, or economic.

The countless combinations of these variables result in countless possible bad outcomes for consumers. To exemplify our thinking, here is one commonly considered scenario: A hacker (actor) breaks into a tracking company (means of access) and publishes its tracking information (action), causing some embarrassing fact about the consumer to become known and inflicting emotional distress (harm).[9]

Risks associated with third-party tracking are heightened by the lack of market pressure to exercise good security and privacy practices. If a first-party website is untrustworthy, users may decline to visit it. But, since users are unaware of the very existence of many third-party websites, they cannot reward responsible sites and penalize irresponsible sites.[10]

D. User Preferences

User surveys have consistently shown opposition to third parties collecting and using browsing activity. A 2009 representative U.S. phone survey by Turow et al. [33] found that 87% of respondents would not want advertising based on tracking. In an unrepresentative 2010 survey of Amazon Mechanical Turk users by McDonald and Cranor [34], only 45% of respondents wanted to be shown any ads that had been tailored to their interests. A December 2010 USA Today/Gallup poll [35] reported 67% of respondents thought behavioral targeting should be outright *illegal*. In a mid-2011 representative U.S. online survey by TRUSTe and Harris Interactive [36], 85% of respondents said they would not consent to tracking for ad targeting, and 78% said they would not consent to tracking for website analytics.

[9]There has not yet been a reported data breach that involved release of third-party web tracking data. (Current data breach notification laws may not extend to third-party web tracking information.) Hackers have begun to target marketing companies; one of the largest data breaches of 2011 was at Epsilon, an email marketing company [32].

[10]Publishers could somewhat stand in for users by demanding good corporate practices, but they have in large measure declined to do so.

Finally, a 2012 representative telephone survey by Pew Research found that 68% of respondents were "not okay" with behavioral advertising [37].

One area for future survey work is in disaggregating user preferences about collection of tracking data from preferences about specific uses of tracking data. The survey literature has largely (but not entirely) focused on behavioral advertising, which can conflate data collection and use.

Another area for future research is preference balancing. All of the above studies examined user preferences independent of economic considerations; there remains a need for work that more directly examines the economic tradeoffs users would make to be or not be tracked.

E. Policy Views

Policy views on third-party web tracking vary substantially. All stakeholders agree that consumers should have some degree of control over web tracking, but there are many points of disagreement on specifics.

- *What should consumers be able to control?* Many policymakers and advocates believe consumers should have control over the *collection* of web tracking information. Online advertising trade groups have argued that control should only extend to specific *uses* of data.
- *What should the default be?* EU policymakers believe no tracking should be the default [38]; advertising trade groups have argued tracking should be the default [39].
- *Who should design the choice mechanism?* Advertising trade groups would like to control choice mechanism design [39]. Many policymakers and advocates believe the browser vendors should retain design responsibility.

Views on web tracking policy are, of course, shaded by underlying priorities. Some, particularly consumer advocates and EU policymakers, view online privacy as a *fundamental human right*. Others, including many researchers and U.S. policymakers, see consumer choice about tracking privacy risks as a means to maximize *welfare*.[11] Mozilla [40], [41] takes the position that giving consumers a *choice* about tracking is itself a policy goal. Third-party websites and advertising trade groups largely defend current practices with arguments rooted in welfare—that the subsidy to content outweighs consumer privacy risks—and *economic rights*.

IV. Regulation and Self-Regulation

Third-party web tracking has, until recently, largely existed in a regulatory vacuum. The following subsections detail limits imposed by U.S. and EU law, as well as the online advertising industry's self-regulatory programs.

[11]We, for example, believe web tracking policy should aim to maximize welfare by setting a default that maximally satisfies aggregate user and website preferences and enabling bargaining with minimal transaction costs.

A. United States

The Federal Trade Commission (FTC) is the leading federal regulatory and law enforcement agency for consumer protection. The FTC has narrowly circumscribed general statutory authority: it can only prevent business practices that are either "unfair" or "deceptive" under 15 U.S.C. § 45. On tracking issues the agency has generally relied on its deception authority, where a company breaches an express representation it has made to consumers.[12] The FTC almost always settles a company's first violation with a consent order and slight (if any) payment; though not directly financially punitive, business are loath to endure the expense, burden, and negative publicity of a federal law enforcement action. A subsequent violation of a consent order can result in significant monetary penalties.

Signaling its heightened interest in the area, the FTC brought three enforcement actions related to third-party web tracking in 2011.

- Chitika, a display advertising network, offered an opt-out cookie that expired after ten days [42].
- ScanScout, an in-video advertising network, used "Flash cookies" but told users they could prevent tracking by disabling cookies [43].
- Facebook claimed that it would not share personal information about its users with advertisers, but it leaked user IDs in referrers for ad clicks and third-party applications [44].

The FTC wields significant soft power that complements its enforcement activity. The agency can threaten enforcement, propose legislation, or publicly call on businesses to improve their practices. The FTC has been particularly vocal on web tracking; since late 2010 [45] commissioners and staff have consistently called for a Do Not Track consumer choice mechanism that is universal, usable, persistent, enforceable, and limits data collection [46] (see Section IX-C).

State attorneys general have consumer protection authority that largely parallels (and in some states exceeds) the FTC's. No attorney general's office has yet brought an enforcement action over tracking-related practices.[13]

Civil class action attorneys have attempted to raise a number of federal and state claims over third-party web tracking practices. In early litigation, several companies agreed to multi-million dollar settlements (e.g. [48]). Defendants in many recent suits have won dismissal on insufficient showing of harm (e.g. [49]).

[12]The FTC has used its unfairness authority in other privacy contexts. Enforcement actions for inadequate security precautions that allowed a data breach, for example, have rested on unfairness.

[13]State attorneys general have been increasingly scrutinizing online privacy practices. The Attorney General of California, for example, recently threatened litigation against mobile application developers that do not provide a privacy policy as required by the Online Privacy Protection Act, Cal. Bus. & Prof. Code §§ 22575-22579 [47].

In early 2012 the White House released a long-awaited online privacy report from a policy collaboration with the Department of Commerce [50]. The report calls for baseline privacy legislation and Commerce-mediated multi-stakeholder codes of conduct that are ratified and enforced by the FTC. The White House and Commerce Department have not indicated their proposals would alter the FTC's present leadership on web tracking issues, and the Chairman of the FTC has suggested he shares that view [51].

B. European Union

The 2002 ePrivacy Directive, 2002/58/EC, mandated that websites enable users to opt out of having information stored in their browser, except as "strictly necessary" to provide service "explicitly requested" by the user. In practice the directive has had little effect; member states have not taken any measures to enforce compliance, and in many cases they have treated browser cookie settings as adequate implementation (see [52]).

A 2009 amendment to the ePrivacy Directive, 2009/136/EC, replaced the opt-out rule with an opt-in consent rule (see [53], [54], [55]). Member state implementations initially split. Some states suggested existing browser cookie settings would remain adequate, on the legal theory that they convey "implicit consent." The majority view, and the developing consensus, is that the directive requires explicit, affirmative consent for each third party, and that Do Not Track (see Section IX-C) could satisfy the consent requirement of the directive. This view has been endorsed by leaders in both the European Commission [56], [57], [58], the EU's executive branch, and the Article 29 Working Party [53], [52], [38], a data protection advisory body. EU and state authorities have yet to enforce compliance with the amended ePrivacy Directive.

In February 2012 the European Commission proposed a new set of revisions to EU data protection law [59]. Recommended provisions would clarify that consent must be explicit, unambiguously extend the reach of regulations to non-EU companies that track EU residents, and impose a stringent penalty structure reaching up to 2% of revenue.

C. Online Advertising Self-Regulation

The online advertising industry has largely harmonized self-regulatory efforts in the U.S. (the Network Advertising Initiative, NAI [60] and the Digital Advertising Alliance, DAA [61]) and the EU (the Interactive Advertising Bureau Europe, IAB Europe [62]). All three programs impose the same consumer choice requirement: participating companies must allow users to opt out of behavioral advertising, that is, ad targeting based on tracking. Note that this is a choice about one particular use of data; collection and other uses of third-party tracking data are unaffected.[14]

Participation in self-regulation has fluctuated with waxing and waning government scrutiny [65]. At present most of the largest online advertising and analytics companies participate, and most of the smaller ones do not. Social networks and content providers are almost entirely absent.

The DAA announced in late 2011 [63] that it would attempt to expand its program to non-advertising businesses and that it would broaden its consumer choice requirement to nearly all uses of third-party data for per-device[15] personalization. Most of the largest social networks and content providers were not stakeholders in the DAA's program expansion and have not signaled acceptance.

There has been scant industry enforcement against businesses that violate self-regulatory principles. In late 2011 the Better Business Bureau announced its first "decisions" against companies that had defective opt-out cookie mechanisms (see Section IX-A); the companies fixed their opt-out cookies, but were not otherwise penalized [66]. The NAI has released an annual "Compliance Report" since 2009 [67], [68], [69]. Only one company has been penalized for non-compliance; it is required to undergo an annual independent privacy audit for three years.

V. Business Models and Trends

There are, broadly, six common business models for third-party websites: advertising companies, analytics services, social networks, content providers, frontend services, and hosting platforms. This taxonomy is intended to assist researchers in modeling third-party businesses; in practice, many services cut across business models, and new business models are frequently attempted.

A. Advertising Companies

While pricing models in online advertising converged by the early 2000s on a small set of auction algorithms (see [70], [71]), marketplace structures vary. There are three main models: direct buy, ad networks, and ad exchanges.

1) Direct Buy: In the oldest model of online advertising, advertisers (and agencies) cut deals directly with first-party websites ("publishers"). This approach fell into disfavor for most websites in the late 1990s through 2000s, but remains the dominant model for search engine and social network advertising. Direct buy has, of late, experienced a renaissance among content publishers owing to the development of "private advertising exchanges," real-time advertising auctions run by publishers. Many implementations of direct buy

[14]The programs impose similar baseline requirements. All three mandate a modest degree of notice and transparency about behavioral advertising, reasonable security precautions for behavioral advertising data, and user consent for behavioral advertising use of narrow classes of sensitive information. All three also prohibit behavioral targeting specifically directed towards children. A recent revision of the DAA principles [63], [64] prohibits certain particularly sensitive uses of information.

[15]The DAA has left the door open to per-user content tailoring, such as personalized social networking widgets [64].

advertising, especially search and social network advertising, do not load content from third-party websites, and therefore do not raise tracking privacy concerns.[16]

2) Advertising Networks: By the late 1990s growth in advertiser demand and ad slot supply ("inventory") made it impractical for advertisers and publishers to deal directly. Ad networks offered a solution by enabling advertisers to easily place ads with many publishers, and by allowing publishers to support their content with many advertisers—with no ad sales team. Networks also brought the ability to systematically target ads to users, based on a publisher's estimated audience ("demographic targeting"), a user's location ("geographic targeting"), a web page's content ("contextual targeting"), or a user's browsing history ("behavioral targeting"). Ad networks remain the largest and most widely used intermediaries in online advertising.

3) Advertising Exchanges: In the mid-2000s publishers began seeking ways to monetize the "remnant" inventory they were not able to sell through an ad network. Ad exchanges offered to fill the slots, in real time, taking bids from many advertisers via many advertising networks ("real-time bidding" or "RTB"). Ad exchanges quickly extended beyond remnants, and a number of intermediary business models now exist in the exchange ecosystem.

- Demand-side platforms (DSPs), which are replacing ad networks as "virtual on-ramps" for advertisers to place bids in multiple ad exchanges.
- Supply-side platforms (SSPs) and yield optimizers, which assist publishers in strategically making inventory offerings to networks and exchanges so as to maximize revenue.
- Data providers, which sell ad targeting data to advertisers in real time. Data providers often base their targeting recommendations on tracking (e.g. Quantcast), information purchased from publishers (e.g. BlueKai), and offline consumer databases (e.g. Datalogix).

B. Analytics Services

Third-party analytics services provide tools for websites to better understand their visitors, including demographics, user agents, and content views and interactions. While implementations of analytics can differ significantly, nearly all services have adopted one of two business models. Some firms (e.g. Adobe) offer analytics as a paid service; they disclaim any legal right to access a client's analytics data except as directed, and they take technical and business precautions to silo data between clients (see Section VIII-B). Other companies offer a free analytics service; they monetize the data they collect by using it for ad targeting (e.g. Quantcast), market understanding (e.g. Google Analytics), and other valuable ends.

[16]There can be non-tracking privacy issues associated with advertising. Microtargeting, for example, may allow an advertiser to draw sensitive inferences about users who click an ad [72].

C. Social Integration

Social integration enables websites to offer personalized content and single sign-on to social network users. The best-known forms of social integration are provided by first-party social networks, most prominently the Facebook like and comment widgets, the Twitter tweet and status widgets, and the Google +1 button. These social networks offer their widgets for free to increase user engagement and conduct market research; there has been some discussion of using social network data for third-party ad targeting [73].

Some social services, such as Disqus, exist almost exclusively in a third-party context. These services tend to operate on a freemium business model, offering more advanced functionality to paying website customers.

Various forms of intermediaries have sprung up to assist websites with social integration. One common business model is social sharing aggregation. Services like AddThis, ShareThis, and Meebo offer free widgets to websites that enable users to share with dozens of social networks. To monetize their widgets, the services collect tracking and usage data and sell it for ad targeting and market research.

Another growing intermediary business model is single sign-on aggregation. Gigya, for example, facilitates single sign-on with many identity providers.

D. Content Providers

Content providers host video, maps, news, weather, stocks, and other media for embedding into websites. Some content providers, including YouTube, offer third-party widgets to both increase user engagement and generate revenue through in-widget advertising. Many others, such as the Associated Press, charge for their content.

E. Frontend Services

Several third parties host JavaScript libraries and APIs that speed webpage loads (e.g. Google Libraries API) and enable new page functionality (e.g. Google Feed API).

F. Hosting Platforms

Some third parties provide services that assist publishers in distributing their own content, such as blog platforms (e.g. Wordpress.com) and content distribution networks (e.g. Akamai).

G. Market Trends

Krishnamurthy and Wills have collected longitudinal web measurements of approximately 1,200 popular websites between 2006 and the present [74], [12], [13]. They report two consistent trends. First, tracking companies are rapidly increasing the share of websites that they span. Large trackers, including Google, Adobe, and Microsoft, have greatly extended their reach through acquisitions. Second, the number of trackers per page is growing rapidly. Websites now frequently embed content from dozens of third parties.

VI. Economics of Third-Party Web Tracking

Proponents of web tracking often make the economic claim that it is needed to subsidize web services through advertising (e.g. [75], [76], [77], [78]). We believe the claim is subject to debate [79], and central questions remain open:

- *Which segments of the online advertising market depend on third-party tracking, and how is it used?* It appears that only a small share of online advertising is behaviorally targeted [79]. The extent to which advertising relies on other uses of tracking is unclear.
- *What marginal tradeoffs do advertisers face for each use of tracking information?* If tracking-based advertising becomes less feasible or more costly, advertisers will reallocate their expenditures.[17] How they choose to reallocate will depend on the effectiveness and cost of the next-best alternatives to tracking-based advertising. Note that effectiveness and cost point in opposite directions—an advertiser may, for example, invest more in an advertising approach that is per-ad slightly less effective but also per-ad significantly less expensive.[18]
- *To what extent can privacy-preserving technologies replace current uses of tracking?* A number of designs have been advanced that, while not perfect substitutes, would enable much of the advertising functionality that tracking supports (see Section VIII-A). Limitations on tracking could incentivize advertising companies to develop and implement privacy-preserving technologies.
- *What proportion of users would consent to tracking or pay if required to access a service?* If diminished tracking-based advertising does impact publishers, they could require visitors to either pay or consent to tracking. Some proportion of users would choose either option rather than forgo the service.

Given the public attention to third-party web tracking, there is surprisingly scant research on these central issues.

A 2009 industry-sponsored paper by Beales [80] has been widely cited (e.g. [50]) for the proposition that behavioral targeting brings in substantially more value than other forms of ad targeting. Beales's study found that behaviorally targeted advertising was roughly twice as expensive and twice as effective as untargeted ("run of network") advertising. There are at least three problems with the methodology used in the study. First, the paper relies on data from a small, unrepresentative sample of advertising networks. Some statistics rely on data from fewer than five companies. The participating companies self-selected and were aware of the purpose of the study. Second, the paper compares behavioral advertising to untargeted advertising. As noted earlier, the relevant comparison is to the next-best alternative (e.g. contextual targeting). Third, the study concludes that behavioral advertising brings value to publishers through increased effectiveness and price. But, as noted earlier, increased price decreases the marginal value of behavioral advertising to advertisers.

Proponents of third-party web tracking have also frequently cited a 2011 paper by Goldfarb and Tucker [81], [82] reporting a 65% decrease in EU advertising effectiveness after the 2002 ePrivacy Directive was transposed by member states. We find four flaws in the Goldfarb and Tucker study. First, the analysis relies exclusively on self-reported data from one company's surveys of web users. The paper does not explain how the data was collected, let alone demonstrate how it is valid and reliable. In fact, the survey data appears to have a number of oddities. It suggests, for example, that after the EU ePrivacy Directive non-EU advertising was twice as effective on EU viewers as on non-EU viewers.

Second, the Goldfarb and Tucker data is not controlled for types of ad targeting. Behavioral advertising may only account for a slight share of the advertising in the study.

Third, the Goldfarb and Tucker study appears to incorrectly assume that the 2002 EU ePrivacy Directive significantly altered online advertising behavior in Europe. In fact, advertising practices in the EU were largely unaffected by the ePrivacy Directive (see Section IV-C).

Fourth, the study seems to overlook changes in the online advertising market. Behavioral advertising was scarce in 2001 and a very small share of online advertising in 2008 [79]. The same time period yielded significant advances in contextual and demographic ad targeting. If the EU law negatively affected behavioral advertising, we should expect an across-the-board performance lift for EU and non-EU ads, with a slightly greater rise in non-EU performance. Instead, the authors predict and demonstrate a significant decrease in EU performance and near-constant non-EU performance.

A final study, by Yan et al. [83], has been widely miscited by supporters of third-party tracking. In that paper, the authors persuasively demonstrate that ideal behavioral targeting could substantially improve the effectiveness of first-party advertising on the Bing search engine. The paper does not examine behavioral advertising in practice or third-party behavioral advertising.

Third-Party Web Tracking Technology

VII. Tracking Technologies

While the debates surrounding web tracking tend to focus on HTTP cookies, there are myriad stateful ("supercookie") and stateless ("fingerprinting") technologies that can be used to pseudonymously correlate web activities.[19]

[17] In economic terms: there are cross-demand elasticities between tracking and non-tracking forms of advertising.

[18] Advertising auction mechanisms further complicate the inquiry, since they limit the surplus that advertisers can capture from better ad targeting.

[19] A note on jargon: when a non-cookie tracking technology is used to recreate a deleted tracking cookie, it is dubbed a "zombie cookie."

Table III
NON-COOKIE WEB TRACKING TECHNOLOGIES

(a) "Supercookies"

HTTP authentication† [84]
HTTP caching ("cache cookies")
cache control
ETags* ("ETag cookies") [85]
Last-Modified [85] (e.g. [86])
cache content
resource (e.g. JavaScript, HTML, CSS, or media)*
status code
redirect location (e.g. [87])
hits and misses (e.g. [88])
TLS/SSL session ID [89]
browsing history††
userData storage (Internet Explorer only)*
HTML5 storage (session, local, and global)*
HTML5 protocol handlers†
HTML5 content handlers†
W3C geolocation API permission†
`window.name` property* (session only)
HTTP strict transport security [90]
plug-in storage* (e.g. Flash local shared objects, or "Flash cookies")
DNS cache

∗ Observed in use by a third-party website.
† User intervention required.
†† Largely inaccessible in newer browsers, but see [88], [91].

(b) Active "Fingerprinting"

operating system
CPU type
user agent
time zone
clock skew
display settings
installed fonts
installed plugins
enabled plugins
supported MIME types
cookies enabled
third-party cookies enabled

(c) Passive "Fingerprinting"

IP address
operating system
user agent
language
HTTP accept headers

A. Stateful Tracking ("Supercookies")

A website can encode a globally unique pseudonymous device identifier into any stateful web technology so long as it persists at least $log_2 n$ bits, where n is the number of Internet-connected devices (presently roughly 5 billion, requiring 33 bits). Table III(a) provides a list of commonly deployed stateful web technologies and notes which have been observed in use for third-party web tracking. The evercookie library [92] provides a reference implementation for many of these tracking techniques.

Soltani et al. [14], McDonald and Cranor [93], and Ayenson et al. [8] report extensive use of Flash storage by popular websites, and Ayenson et al. found some use HTML5 local storage.

A number of online advertising companies, including ClearSpring, Interclick, Specific Media, and Quantcast, have been discovered using Flash cookies to track users. In mid-2011 Soltani [9] found that a third-party analytics service, KISSmetrics, was using cookies, Flash cookies, ETag cookies, cache cookies, userData, and HTML5 local storage, and that the non-cookie tracking technologies were used to recreate a cookie if deleted. We discovered that Microsoft was using an ETag cookie and a cache cookie in connection with its script for syncing an advertising identifier across web properties [94].

B. Stateless Tracking ("Fingerprinting")

A website may be able to learn properties about the browser that, taken together, form a unique or nearly unique identifier [95], [96]. Some properties require active discovery through a script or plug-in (Table III(b)). Other properties can be passively learned from network traffic (Table III(c)).

In a 2010 sample of nearly 500,000 browsers Eckersley reported 83.6% were uniquely identified with a subset of active fingerprinting features. 94.2% of browsers with Flash or Java enabled were uniquely identified. While fingerprints changed quickly, a simple matching algorithm was able to associate new and old fingerprints with over 99% precision.

Several companies, including 41st Parameter/AdTruth, BlueCava, and iovation, advertise commercial browser fingerprinting technology.

Passive fingerprinting is particularly problematic since it cannot be detected with web measurement. Further research is needed to understand how effective passive fingerprinting is and what steps websites can take to scrub passive fingerprinting data from their logs. A recent study of Hotmail and Bing users by Yen et al. [97] suggests passive fingerprinting may be sufficient to track many stationary browsers.

VIII. PRIVACY-PRESERVING THIRD-PARTY SERVICES

There have been several efforts at designing third-party services that would capture the economic value of particular uses of tracking while preserving user privacy. Current proposals are based on a narrow subset of business models; further work is needed to support privacy across the range of evolving third-party website business models (Section V).

A. Behavioral Advertising

Privad [98] is designed to conceal a user's activities from an advertising network by interposing an anonymizing proxy between the browser and the ad network. In this approach, trusted client software subscribes to streams of possibly relevant ads, selects relevant ads locally, submits candidates for auction, and then reports results. While the Privad model is designed to offer comprehensive privacy guarantees, it requires broad adoption of high-performance anonymizing proxies. This seems unlikely in the near future.

In an extension to Privad [99], Reznichenko et al. evaluate designs for privacy-preserving advertising auctions. The work emphasizes the trade-off between an advertising company's ability to conceal its ranking algorithm and bids and a user's ability to prevent pseudonymous tracking.

Like Privad, Adnostic [100] uses client-based functionality to perform ad selection, but it eliminates anonymizing proxies at the cost of less precise ad targeting. Adnostic also simplifies cost-per-click billing by allowing the advertising network to learn of a user's ad clicks. Cost-per-impression billing would still require a low-performance trusted intermediary so as to not reveal the user's ad impressions. As implemented, Adnostic requires a browser extension, which is a practical barrier to more widespread adoption.

RePriv [101], by Fredrikson and Livshits, is a verifiable policy architecture that enables users to selectively grant permission for generating and sharing client-side data stores that enable website personalization. The RePriv model holds promise as a general-purpose platform for building privacy-preserving advertising like Privad and Adnostic. But, like Adnostic, RePriv would have to be translated from its current implementation as a single-platform browser extension into existing web technologies for near-term deployment.

Bilenko and Richardson [102] propose an approach for keyword-based search advertising that provides privacy against a weaker threat model. The search advertising company is trusted to temporarily compute on user profile data, but then store the data in the browser and delete its copy. The authors ran their algorithm against 60 days of Bing search advertising logs and achieved almost all the benefit of current server-side behavioral targeting. Specifically, they report capturing over 95% of the increase in click-through rates, generating approximately 4% greater revenue than search advertising without behavioral targeting. We are skeptical that the temporary data-use model is likely to be adopted; web services in general, and online advertising companies in particular, have historically been loath to voluntarily discard logs. The model also introduces the risk of inadvertent or surreptitious collection of third-party tracking data.

B. Analytics

Some analytics services have taken technical and legal precautions to silo tracking data for each first-party website.

Several free and paid services, including Google Analytics and Adobe SiteCatalyst (formerly Omniture), use the same-origin policy to restrict the scope of pseudonymous identifiers to a first-party website. Google uses a first-party cookie to achieve this; Adobe offers the choice of a cookie scoped to a first-party subdomain CNAMEd to Adobe (e.g. `metrics.apple.com`) or a cookie scoped to a unique Adobe subdomain (e.g. `paypal.112.2o7.net`).

Google Analytics offers an opt-in feature to websites that prevents logging the last octet of a user's IP address (`anonymizeIp`).[20] This privacy option does not seem to reduce the benefit of the service since Google Analytics does not report IP addresses, and geolocation (the only reported measurement that relies on IP addresses) is unlikely to vary much by the last octet. We nonetheless found barely any use of the option: in an August 2011 crawl of the Alexa top 10,000 global websites, `anonymizeIp` was set on only 63 of 4861 (1.3%) reports to Google Analytics.

Paid analytics services usually promise by contract to make no use of the data they collect except as directed by their clients, and they impose internal business controls to ensure each client's data remains segregated. Adobe, for example, makes these guarantees [103]: "Although the data generated by Adobe's products resides on Adobe's servers, each customer owns the data generated by the use of its site. By contract, Adobe has no right to access or use this data. In addition, Adobe does not allow use of the data for any purpose other than those of the owner (web publisher); that is, Adobe silos each customer's data for use by that customer."

[20] It is unclear how much privacy is afforded by this measure [97].

IX. User Choice Mechanisms

Three technical solutions have been advanced for giving users control over third-party web tracking: opt-out cookies, blocking, and Do Not Track.

A. Opt-Out Cookies and the AdChoices Icon

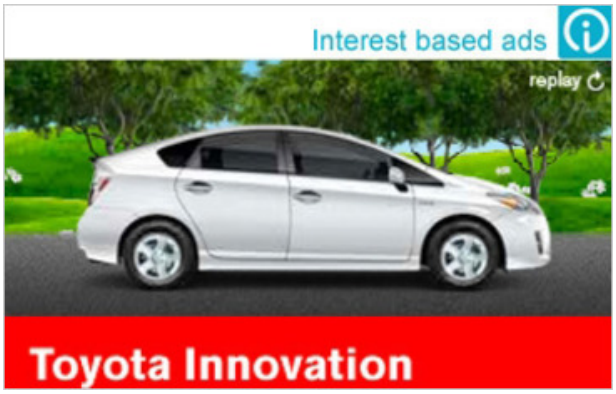
User choice in current online advertising self-regulation is implemented with opt-out cookies. There are several problems with this approach. First, it requires manual updating. To opt out of new third parties, a user has to install new cookies. Second, cookies expire, so a user has to periodically renew opt-out cookies. Third, users may clear their cookies, inadvertently removing their opt-out preferences. Fourth, opt-out cookies are fragile; it is easy for a third party to improperly set or delete an opt-out cookie. Fifth, opt-out cookies scale poorly; each third-party PS+1 requires a network roundtrip, resulting in a sluggish user experience when changing many preferences. Browser extensions for persisting opt-out cookies, such as TACO or Google Keep My Opt Outs, largely mitigate these issues at the cost of usability.

Many online advertising companies have begun to insert an "AdChoices" icon (13x13px) and text (10pt) into display ads (Figure 2(b)) to increase user awareness of behavioral targeting and existing self-regulatory choice mechanisms. Clicking the icon provides additional information about how the ad was targeted and, in many cases, a link to landing page where the user can set opt-out cookies.

Several studies have called into question the usability of the self-regulatory opt-out model.

Before the deployment of the AdChoices icon an industry-funded policy group conducted a large-audience usability survey [106]. It found that a 31x31px icon with 18pt font (Figure 3(a)) was not very effective at conveying information about behavioral targeting practices ("substantial repetition and consumer education may be needed to improve [the

Figure 2. Evolution of the AdChoices icon.



(a) Proposed icon and text [104] (actual size at 115 DPI).

(b) Implemented icon and text [105] (actual size at 115 DPI).

icon's] communication effectiveness over time"), and that the text "AdChoice" performed worse than alternatives.

McDonald and Cranor [34] conducted a large-audience survey on user perceptions of a self-regulatory opt-out page. 88% of participants understood that the page related to online advertising and opting out, but only 11% correctly responded that the page allows opting out of behavioral targeting, not tracking (34%), ads from specific companies (25%), or some proportion of advertising (18%).

Leon et al. [107] examined the usability of two other self-regulatory websites with five in-laboratory participants each. On one website, the Digital Advertising Alliance, only one of the five users was able to opt out without guidance, and none of the users correctly understood the implications of opting out. On the other website, Evidon, four of the five users were able to opt out without guidance, though it took the participant who chose to opt out of all companies 47 minutes to exercise his or her preferences. Once again none of the users correctly understood what opting out would do.

Leon et al. also studied the usability of the TACO extension. All five participants enabled persistent opt-out cookies, the default setting.

Hernandez et al. [105] measured the prevalence of the AdChoices program on the Alexa U.S. 500 top homepages. They found an icon in only 9.9% of third-party ads and an icon and text in only 5.1%.

The online advertising trade groups have declined to provide overall usage statistics of opt-out cookies. Anecdotal reports (e.g. [108]) place usage at less than 1% of browsers.

B. Blocking

Given the myriad approaches to tracking a browser—many of which require nothing more than an HTTP round-trip—the most effective user self-help tools[21] function by blocking third-party web content. Nearly all tools consist of a block list, either available as a subscription for a browser extension or wrapped in a configurable browser extension.

To understand the effectiveness of blocking, we conducted three consecutive FourthParty crawls of the Alexa U.S. top 500 with each of 11 blocking tools installed [110]. We also conducted a baseline crawl to estimate which PS+1s were third-party trackers. For each tool we calculated three values relative to baseline and averaged across all trackers: pages with an HTTP request to a tracker, pages with an HTTP Set-Cookie response from a tracker, and cookies added less cookies deleted by a tracker.[22] Figure 3 presents our results.

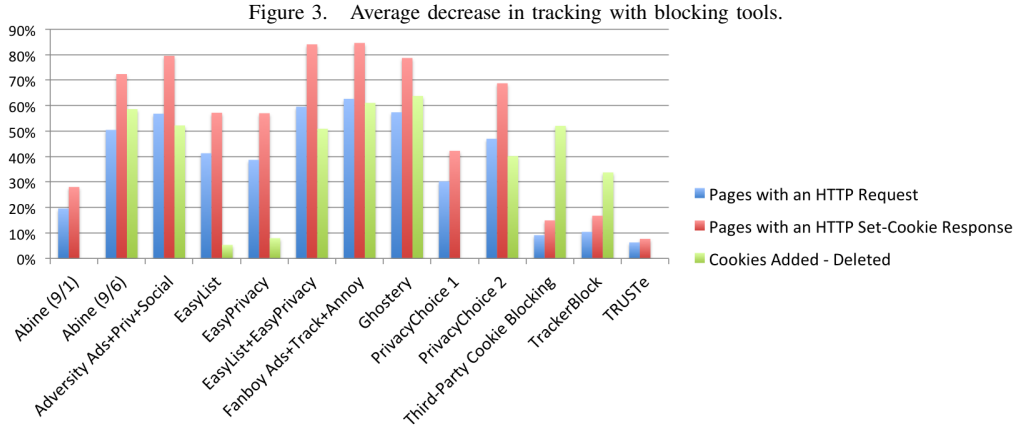
We found significant variability in performance. The most effective tool was a combination of community-maintained Fanboy's Lists for blocking ads, surreptitious tracking, and social content. All of the top performing tools blocked third-party advertising, an unsurprising result since there is no clean division between advertising content and advertising-related tracking content. The block list from TRUSTe was not only the least effective, but it also would override other lists to allow tracking by several sizable third parties.

Leon et al. [107] examined the usability of the Ghostery, Adblock Plus, and Internet Explorer Tracking Protection List blocking tools. Two of the five Ghostery users believed they had enabled the extension's blocking feature when, in fact, they had not. All five of the Adblock Plus users configured the extension with a default advertisement blocking list; none installed additional lists to block non-advertising trackers. All five Internet Explorer Tracking Protection List users

[21] For brevity we do not address private browsing [109], third-party cookie blocking, and other self-help approaches to mitigating tracking. A fuller discussion is available in [110].

[22] We included this noisy metric to roughly gauge the effectiveness of third-party cookie blocking and TrackerBlock, a tool that prevents several forms of stateful tracking.

Figure 3. Average decrease in tracking with blocking tools.



retained the default setting, to not block any content; they all believed they had configured the option to substantially or completely block tracking.

In sum: blocking can be fairly effective, but it is only a realistic solution for advanced users.

C. Do Not Track

Do Not Track uses a combination of technology and policy to provide consumer choice over web tracking. The World Wide Web Consortium (W3C) is presently standardizing Do Not Track; the W3C's working group has not yet reached consensus on the technology or policy components.

The Do Not Track technology is simply an HTTP header, `DNT`, that signals a user's preference about web tracking. Associated technologies have been proposed that would allow a website to request exceptions and signal its own tracking status. Firefox, Internet Explorer, Safari, and Opera presently support a Do Not Track opt-out preference (sending the `DNT: 1` header). Google has pledged to add the feature to Chrome. As of late 2011, Mozilla [111] reported 5.6% usage in desktop Firefox and 17.1% usage in Firefox Mobile.

Roughly twenty websites presently honor the Do Not Track technology, and the Digital Advertising Alliance recently pledged [39] that its about eighty member companies would begin supporting the header.

Do Not Track enforcement could be accomplished through measurement of tracking technologies, using tools like FourthParty.[23] In mid-2011 we identified two advertising companies that were surreptitiously taking steps to honor Do Not Track [113], suggesting the approach is quite viable.

The Do Not Track policy defines what websites must do when they receive a Do Not Track header. Debates over the Do Not Track policy have been largely coextensive with debates over third-party web tracking policy (see Section III-E). Policymakers, consumer advocates, and researchers are in general agreement that Do Not Track must significantly curtail third-party information collection. The recent DAA commitment only requires a third-party website to stop per-device content personalization if it receives a Do Not Track signal (see Section IV-C).

[23] In the advertising space, Do Not Track might also be enforced by monitoring ad distributions for evidence of behavioral targeting. It is unclear how feasible this approach is [112].

X. Conclusion

This paper surveyed policy and technology issues in third-party web tracking as of early 2012. The field is rapidly changing; new announcements, questions, and research results appear by the week. We hope the information presented here provides security and privacy researchers with the background necessary to contribute to this developing field and to meaningfully participate in the ongoing public debate.

Acknowledgements

We thank Jovanni Hernandez and Akshay Jagadeesh for their invaluable research assistance. This paper benefited from feedback provided by Nick Doty, Peter Eckersley, Aleecia McDonald, Hart Montgomery, Arvind Narayanan, Ashkan Soltani, Thomas Steinke, and many others. All errors and omissions are solely our own.

The authors acknowledge the support of the National Science Foundation, the Air Force Office of Scientific Research, the Office of Naval Research, and Stanford University.

References

[1] World Wide Web Consortium. Content Security Policy. [Online]. Available: http://w3.org/TR/CSP/

[2] J. Grossman, R. Hansen, P. D. Petkov, A. Rager, and S. Fogie, *XSS Attacks: Cross-Site Scripting Exploits and Defense*. Burlington, MA: Syngress, 2007.

[3] A. Barth, C. Jackson, and J. C. Mitchell, "Robust defenses for cross-site request forgery," in *Proceedings of the 2008 ACM Conference on Computer and Communications Security*, October 2008.

[4] W. Zeller and E. W. Felten, "Cross-site request forgeries: Exploitation and prevention," Princeton University, Tech. Rep., September 2008.

[5] W. Enck, P. Gilbert, B. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones," in *Proceedings of the 9th USENIX Symposium on Operating Systems Design and Implementation*, October 2010.

[6] S. Thurm and Y. I. Kane, "Your apps are watching you," *The Wall Street Journal*, December 2010.

[7] J. Mayer. (2011, July) Tracking the trackers: To catch a history thief. [Online]. Available: http://cyberlaw.stanford.edu/node/6695

[8] M. Ayenson, D. J. Wambach, A. Soltani, N. Good, and C. J. Hoofnagle, "Flash cookies and privacy II: Now with HTML5 and ETag respawning," July 2011.

[9] A. Soltani. (2011, August) Respawn redux. [Online]. Available: http://ashkansoltani.org/docs/respawn_redux.html

[10] B. Krishnamurthy and C. Wills, "Privacy leakage vs. protection measures: the growing disconnect," in *Proceedings of the Web 2.0 Security and Privacy Workshop*, May 2011.

[11] ——, "On the leakage of personally identifiable information via online social networks," in *Proceedings of the ACM Workshop on Online Social Networks*, August 2009.

[12] B. Krishnamurthy and C. E. Wills, "Privacy diffusion on the web: A longitudinal perspective," in *Proceedings of the 18th Conference on the World Wide Web*, April 2009.

[13] ——, "Generating a privacy footprint on the Internet," in *Proceedings of the 6th ACM Conference on Internet Measurement*, October 2006.

[14] A. Soltani, S. Canty, Q. Mayo, L. Thoma, and C. J. Hoofnagle, "Flash cookies and privacy," August 2009.

[15] F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web," in *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation*, April 2012.

[16] P. G. Leon, L. F. Cranor, A. M. McDonald, and R. McGuire, "Token attempt: The misrepresentation of website privacy policies through the misuse of P3P compact policy tokens," in *Proceedings of the 2010 Workshop on Privacy in the Electronic Society*, October 2010.

[17] D. Jang, R. Jhala, S. Lerner, and H. Shacham, "An empirical study of privacy-violating information flows in JavaScript web applications," in *Proceedings of the 2006 ACM Conference on Computer and Communications Security*, October 2010.

[18] ECMA. Harmony proxies. [Online]. Available: http://wiki.ecmascript.org/doku.php?id=harmony:proxies

[19] J. Mayer. (2011, October) Tracking the trackers: Where everybody knows your username. [Online]. Available: http://cyberlaw.stanford.edu/node/6740

[20] A. Narayanan. (2011, July) There is no such thing as anonymous online tracking. [Online]. Available: http://cyberlaw.stanford.edu/node/6701

[21] Datalogix. Datalogix privacy policy. [Online]. Available: http://datalogix.com/privacy/

[22] D. Perito, C. Castelluccia, M. A. Kaafar, and P. Manilsr, "How unique and traceable are usernames?" in *Proceedings of the 2011 Privacy Enhancing Technologies Symposium*, 2011.

[23] D. Irani, S. Webb, C. Pu, and K. Li, "Personal-information leakage from multiple online social networks," *IEEE Internet Computing*, May 2011.

[24] D. Irani, S. Webb, K. Li, and C. Pu, "Large online social footprints - an emerging threat," in *Proceedings of the 2009 International Conference on Computational Science and Engineering*, August 2009.

[25] A. Narayanan. (2008, November) Lendingclub.com: A de-anonymization walkthrough. [Online]. Available: http://33bits.org/2008/11/12/57/

[26] Mozilla Foundation. Public suffix list. [Online]. Available: http://publicsuffix.org/

[27] A. Narayanan. (2010) How Google Docs leaks your identity. [Online]. Available: http://33bits.org/2010/02/22/google-docs-leaks-identity/

[28] L.-S. Huang and C. Jackson, "Clickjacking attacks unresolved," July 2011.

[29] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large datasets," in *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, May 2008.

[30] ——, "De-anonymizing social networks," in *Proceedings of the 2009 IEEE Symposium on Security and Privacy*, May 2009.

[31] A. Acquisti, R. Gross, and F. Stutzman, "Faces of Facebook," in *Black Hat 2011*, August 2011.

[32] Epsilon. (2011, April) Epsilon notifies clients of unauthorized entry into email system. [Online]. Available: http://epsilon.com/news-events/press-releases/2011/epsilon-notifies-clients-unauthorized-entry-email-system

[33] J. Turow, J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy, "Americans reject tailored advertising and three activities that enable it," September 2009.

[34] A. M. McDonald and L. F. Cranor, "Beliefs and behaviors: Internet users' understanding of behavioral advertising," in *Proceedings of the 2010 Research Conference on Communication, Information and Internet Policy*, October 2010.

[35] Gallup. (2010, December) USA Today/Gallup poll. [Online]. Available: http://gallup.com/poll/File/145334/Internet_Ads_Dec_21_2010.pdf

[36] TRUSTe and Harris Interactive. (2011, July) Privacy and online behavioral advertising. [Online]. Available: http://truste.com/ad-privacy/TRUSTe-2011-Consumer-Behavioral-Advertising-Survey-Results.pdf

[37] Pew Research Center. (2012, March) Search engine use 2012. [Online]. Available: http://pewinternet.org/Reports/2012/Search-Engine-Use-2012.aspx

[38] Article 29 Data Protection Working Party. (2012, March) Letter to the online advertising industry. [Online]. Available: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2012/20120301_reply_to_iab_easa_en.pdf

[39] Digital Advertising Alliance. (2012, February) DAA position on browser based choice mechanism. [Online]. Available: http://aboutads.info/resource/download/DAA_Committment.pdf

[40] S. Stamm. (2011, November) Why we won't enable DNT by default. [Online]. Available: http://blog.mozilla.com/privacy/2011/11/09/dnt-cannot-be-default/

[41] T. Lowenthal. (2011, November) Deeper discus-

sion of our decision on DNT defaults. [Online]. Available: http://blog.mozilla.com/privacy/2011/11/15/deeper-discussion-of-our-decision-on-dnt-defaults/

[42] Federal Trade Commission. (2011, March) FTC puts an end to tactics of online advertising company that deceived consumers who wanted to "opt out" from targeted ads. [Online]. Available: http://ftc.gov/opa/2011/03/chitika.shtm

[43] ——. (2011, November) Online advertiser settles FTC charges ScanScout deceptively used Flash cookies to track consumers online. [Online]. Available: http://ftc.gov/opa/2011/11/scanscout.shtm

[44] ——. (2011, November) Facebook settles FTC charges that it deceived consumers by failing to keep privacy promises. [Online]. Available: http://ftc.gov/opa/2011/11/privacysettlement.shtm

[45] Federal Trade Commission Staff, "Protecting consumer privacy in an era of rapid change," December 2010. [Online]. Available: http://ftc.gov/os/2010/12/101201privacyreport.pdf

[46] E. Felten, "FTC perspective," presented at W3C Workshop on Web Tracking and User Privacy, April 2011. [Online]. Available: http://w3.org/2011/track-privacy/slides/Felten.pdf

[47] California Department of Justice. (2012, February) Attorney General Kamala D. Harris secures global agreement to strengthen privacy protections for users of mobile applications. [Online]. Available: http://oag.ca.gov/news/press_release?id=2630

[48] J. Mullen, "Judge approves $ 2.4 million Quantcast privacy settlement," *paidContent*, June 2011. [Online]. Available: http://paidcontent.org/article/419-judge-approves-2.4-million-quantcast-privacy-settlement/

[49] ——, "Second privacy lawsuit over 'Flash cookies' falls apart," *paidContent*, August 2011. [Online]. Available: http://paidcontent.org/article/419-privacy-lawsuits-over-flash-cookies-falling-apart/

[50] Executive Office of the President, "Consumer data privacy in a networked world," February 2012. [Online]. Available: http://whitehouse.gov/sites/default/files/privacy-final.pdf

[51] J. Mayer. (2012, February) The FTC's chairman groks Do Not Track. [Online]. Available: http://webpolicy.org/2012/02/29/the-ftcs-chairman-groks-do-not-track/

[52] Article 29 Data Protection Working Party. (2011, August) Letter to the online advertising industry. [Online]. Available: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2011/20110803_letter_to_oba_annexes.pdf

[53] ——. (2010, June) Opinion 2/2010 on online behavioural advertising. [Online]. Available: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf

[54] M. Rasmussen. (2011, September) European Union ePrivacy Directive—latest update. [Online]. Available: http://blogs.omniture.com/2011/09/01/european-union-eprivacy-directive-update/

[55] K. Retzer and J. Lopatowska. (2011, May) How to do cookies without clear directions. [Online]. Available: http://mofo.com/files/Uploads/Images/110516-ePrivacy-Directive.pdf

[56] N. Kroes, "Online privacy - reinforcing trust and confidence," June 2011. [Online]. Available: http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/461

[57] ——, "Why we need a sound Do-NotTrack standard for privacy online," January 2012. [Online]. Available: http://blogs.ec.europa.eu/neelie-kroes/donottrack/

[58] ——, "Privacy online: USA jumps aboard the "Do-Not-Track" standard," February 2012. [Online]. Available: http://blogs.ec.europa.eu/neelie-kroes/usa-do-not-track/

[59] European Commission, "Commission proposes a comprehensive reform of the data protection rules," January 2012. [Online]. Available: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm

[60] Network Advertising Initiative, "2008 NAI principles," 2008. [Online]. Available: http://networkadvertising.org/Principles.pdf

[61] Digital Advertising Alliance, "Self-regulatory principles for online behavioral advertising," July 2009. [Online]. Available: http://aboutads.info/resource/download/seven-principles-07-01-09.pdf

[62] Interactive Advertising Bureau Europe, "IAB Europe EU framework for online behavioural advertising," July 2011. [Online]. Available: http://iabeurope.eu/media/55448/iab%20europe%20report%20july_28.pdf

[63] Digital Advertising Alliance, "Self-regulatory principles for multi-site data," November 2011. [Online]. Available: http://aboutads.info/resource/download/Multi-Site-Data-Principles.pdf

[64] J. Mayer. (2011, November) A brief overview of the supplementary DAA principles. [Online]. Available: http://cyberlaw.stanford.edu/node/6755

[65] R. Gellman and P. Dixon, "Many failures: A brief history of privacy self-regulation in the united states," October 2011. [Online]. Available: http://worldprivacyforum.org/pdf/WPFselfregulationhistory.pdf

[66] R. Reitman. (2011, November) The DAA's self-regulatory principles fall far short of Do Not Track. [Online]. Available: https://eff.org/deeplinks/2011/11/daa-self-regulation-principles-fall-far-short-do-not-track

[67] Network Advertising Initiative, "2011 annual compliance report," 2012. [Online]. Available: http://networkadvertising.org/pdfs/NAI_2011_Compliance_Report.pdf

[68] ——, "2010 annual compliance report," 2011. [Online]. Available: http://networkadvertising.org/pdfs/2010_NAI_Compliance_Report.pdf

[69] ——, "2009 annual compliance report," 2009. [Online]. Available: http://networkadvertising.org/pdfs/2009_NAI_Compliance_Report_12-30-09.pdf

[70] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second-price auction," *The American Economic Review*, vol. 97, no. 1, March 2007.

[71] H. R. Varian, "Position auctions," *International Journal of Industrial Organization*, vol. 25, 2007.

[72] A. Korolova, "Privacy violations using microtargeted ads," *Journal of Privacy and Confidentiality*, 2011.

[73] E. J. Markey and J. Barton, "Letter to Mark Zuckerberg," November 2011.

[74] B. Krishnamurthy, "Privacy leakage on the Internet," presented at IETF 77, March 2010. [Online]. Available: http://ietf.org/proceedings/77/slides/plenaryt-5.pdf

[75] T. Vega and V. Kopytoff, "In online privacy plan, the opt-out question looms," *The New York Times*, December 2010.

[76] M. Zaneis, "'Do Not Track' rules would put a stop to the

Internet as we know it," *U.S. News*, January 2011, opinion.

[77] B. Kunz, "The $8 billion Do Not Track prize," *Bloomberg Businessweek*, December 2010, opinion.

[78] H. David, "Do Not Track: Revenue impact on online advertising," *Bloomberg Government*, March 2011.

[79] J. Mayer. (2011, January) Do Not Track is no threat to ad-supported businesses. [Online]. Available: http://cyberlaw.stanford.edu/node/6592

[80] H. Beales, "The value of behavioral targeting," March 2010. [Online]. Available: http://networkadvertising.org/pdfs/Beales_NAI_Study.pdf

[81] A. Goldfarb and C. E. Tucker, "Privacy regulation and online advertising," *Management Science*, January 2011.

[82] ——, "Online advertising, behavioral targeting, and privacy," *Communications of the ACM*, May 2011.

[83] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen, "How much can behavioral targeting help online advertising?" in *Proceedings of the 18th Conference on the World Wide Web*, April 2009.

[84] J. Franks, P. Hallam-Baker, J. Hostetler, S. Lawrence, P. Leach, A. Luotonen, and L. Stewart, "HTTP authentication: Basic and digest access authentication," RFC 2617, June 1999.

[85] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, "Hyptertext transfer protocol – HTTP/1.1," RFC 2616, June 1999.

[86] N. Cubrilovic. (2011, August) Persistent and unblockable cookies using HTTP headers. [Online]. Available: http://nikcub.appspot.com/posts/persistant-and-unblockable-cookies-using-http-headers

[87] E. Bursztein. (2011, July) Tracking users that block cookies with a HTTP redirect. [Online]. Available: http://elie.im/blog/security/tracking-users-that-block-cookies-with-a-http-redirect/

[88] M. Zalewski. (2011, December) Rapid history extraction through non-destructive cache timing. [Online]. Available: http://lcamtuf.coredump.cx/cachetime/

[89] T. Dierks and E. Rescorla, "The transport layer security (TLS) protocol version 1.2," RFC 5246, August 2008.

[90] J. Hodges, C. Jackson, and A. Barth, "HTTP strict transport security (HSTS)," draft-ietf-websec-strict-transport-sec-06, March 2012.

[91] Z. Weinberg, E. Chen, P. R. Jayaraman, and C. Jackson, "I still know what you visited last summer: Leaking browsing history via user interaction and side channel attacks," in *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, May 2011.

[92] S. Kamkar. (2010, September) evercookie. [Online]. Available: http://samy.pl/evercookie/

[93] A. M. McDonald and L. F. Cranor, "A survey of the use of Adobe Flash local shared objects to respawn HTTP cookies," Carnegie Mellon CyLab, Tech. Rep. 11-001, January 2011.

[94] J. Mayer. (2011, August) Tracking the trackers: Microsoft advertising. [Online]. Available: http://cyberlaw.stanford.edu/node/6715

[95] ——, "'Any person... a pamphleteer': Internet anonymity in the age of web 2.0," Undergraduate thesis, Princeton University, Princeton, NJ, May 2009.

[96] P. Eckersley, "How unique is your web browser?" in *Proceedings of the 2010 Privacy Enhancing Technologies Symposium*, July 2010.

[97] T.-F. Yen, Y. Xie, F. Yu, R. P. Yu, and M. Abadi, "Host fingerprinting and tracking on the web: Privacy and security implications," in *Proceedings of the 19th Annual Network and Distributed System Security Symposium*, February 2012.

[98] S. Guha, B. Cheng, and P. Francis, "Privad: Practical privacy in online advertising," in *Proceedings of the 2011 USENIX Symposium on Networked Systems Design and Implementation*, April 2011.

[99] A. Reznichenko, S. Guha, and P. Francis, "Auctions in Do-Not-Track compliant Internet advertising," in *Proceedings of the 2011 ACM Conference on Computer and Communications Security*, October 2011.

[100] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas, "Adnostic: Privacy preserving targeted advertising," in *Proceedings of the 2010 Network and Distributed System Security Symposium*, March 2010.

[101] M. Fredrikson and B. Livshits, "Repriv: Re-envisioning in-browser privacy," in *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, May 2011.

[102] M. Bilenko and M. Richardson, "Predictive client-side profiles for personalized advertisingg," in *Proceedings of the 2011 ACM Conference on Knowledge Discovery and Data Mining*, August 2011.

[103] M. J. Rasmussen, "Adobe position paper on privacy and tracking," in *W3C Workshop on Web Tracking and User Privacy*, April 2011.

[104] S. Clifford, "A little 'i' to teach about online privacy," *The New York Times*, January 2010.

[105] J. Hernandez, A. Jagadeesh, and J. Mayer. (2011, August) Tracking the trackers: The AdChoices icon. [Online]. Available: http://cyberlaw.stanford.edu/node/6714

[106] M. Hastak and M. J. Culnan, "Online behavioral advertising icon study," January 2010. [Online]. Available: http://futureofprivacy.org/final_report.pdf

[107] P. G. Leon, B. Ur, R. Balebako, L. F. Cranor, R. Shay, and Y. Wang, "Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising," Carnegie Mellon CyLab, Tech. Rep. 11-017, October 2011.

[108] World Wide Web Consortium. (2011, April) Web tracking and user privacy workshop. [Online]. Available: http://w3.org/2011/04/29-w3cdnt-minutes.html

[109] G. Aggrawal, E. Bursztein, C. Jackson, and D. Boneh, "An analysis of private browsing modes in modern browsers," in *Proceedings of the 19th USENIX Security Symposium*, 2010.

[110] J. Mayer. (2011, September) Tracking the trackers: Self-help tools. [Online]. Available: http://cyberlaw.stanford.edu/node/6730

[111] A. Fowler. (2011, November) Do Not Track Adoption in Firefox Mobile is 3x higher than desktop. [Online]. Available: http://blog.mozilla.com/privacy/2011/11/02/do-not-track-adoption-in-firefox-mobile-is-3x-higher-than-desktop/

[112] S. Guha, B. Cheng, and P. Francis, "Challenges in measuring online advertising systems," in *Proceedings of the 10th ACM Conference on Internet Measurement*, November 2010.

[113] J. Mayer. (2011, July) Tracking the trackers: Early results. [Online]. Available: http://cyberlaw.stanford.edu/node/6694