Web Tracking – A Literature Review on the State of Research

Tatiana Ermakova
University of Potsdam
tatiana.ermakova@uni-potsdam.de

Benedict Bender
University of Potsdam
benedict.bender@wi.uni-potsdam.de

Benjamin Fabian
HfT Leipzig
fabian@hft-leipzig.de

Kerstin Klimek
University of Potsdam
kerstin.klimek@uni-potsdam.de

Abstract

Web tracking seems to become ubiquitous in online business and leads to increased privacy concerns of users. This paper provides an overview over the current state of the art of web-tracking research, aiming to reveal the relevance and methodologies of this research area and creates a foundation for future work. In particular, this study addresses the following research questions: What methods are followed? What results have been achieved so far? What are potential future research areas? For these goals, a structured literature review based upon an established methodological framework is conducted. The identified articles are investigated with respect to the applied research methodologies and the aspects of web tracking they emphasize.

1. Introduction

Compared to digital advertising, where advertisers and publishers sign private deals, programmatic advertising automates the purchase of digital ad-inventory, in a common case by means of real-time auctioning and bidding (Stange & Funk, 2015). Programmatic advertising attracts an increasing number of marketers and advertisers (O'Connell, 2014) by allowing them to reach their target audiences within the "right" context and, hence, generate higher returns on their brand campaigns (Fernandez-Tapia, 2016). Online users' browsing behavior is seen as a worthwhile source for building their detailed profiles (Mitchell, 2012; Falahrastegar et al., 2016), being of high relevance to improve the above outlined commercial activities (Roesner et al., 2012).

Against this background, online users are increasingly tracked in real time and across multiple websites (Gomer et al., 2013, Falahrastegar et al., 2014), although presumably with different levels of intensity (Ermakova et al., 2017), and even by emails (Fabian et al., 2015; Bender et al., 2016). Hence, driven by a variety of enabling techniques (Besson et al., 2014; Sanchez-Rola et al., 2016), web tracking has become ubiquitous on the Web (Roesner et al., 2012), across websites and even across devices (Brookman et al., 2017). Besides targeted advertising (Sanchez-Rola et al., 2016; Parra-Arnau, 2017), web tracking can be employed for personalization (Sanchez-Rola et al., 2016; Mayer & Mitchell, 2012; Roesner et al., 2012), advanced web site analytics, social network integration (Mayer & Mitchell, 2012; Roesner et al., 2012), and website development (Fourie & Bothma, 2007).

For online users, especially mature, well-off and educated individuals, who constitute the most preferred target group of web tracking (Peacock, 2015), the web tracking practices also imply higher privacy losses (Mayer & Mitchell, 2012; Roesner et al., 2012) and risks including price discrimination, government surveillance, and identity theft (Bujlow et al., 2015, 2017). For instance, Narayanan & Shmatikov (2009) could correctly identify over one third of users given their social patterns on Twitter and Flickr.

Despite the popularity of web tracking within commercial and research communities (Libert, 2015; Hamed et al., 2013; Acar et al., 2014; Han et al., 2012; Roesner et al., 2012; Schelter & Kunegis, 2016a, 2016b; Englehardt & Narayanan, 2016; Gomer et al., 2013), earlier works mainly present single aspects of the topic.

As a literature review on the state of research on web tracking, this paper aims to reveal the relevance and methods of this research field (vom Brocke et al., 2009) and creates a foundation for further research (Baker, 2000). In particular, the focus is placed on the following research questions: (1) What methods are followed? (2) What results have been achieved so far? (3) What are potential future research areas?

2. Method

URI: http://hdl.handle.net/10125/50485
ISBN: 978-0-9981331-1-9

For this literature review, we follow a five-step approach by Herz et al. (2010), which requires review scope definition, topic conceptualization, literature search, analysis and synthesis, and research agenda.

2.1. Definition of the review scope

As recommended by vom Brocke et al. (2009), we apply Cooper's (1988) taxonomy for review scope definition. Specifically, we concentrate on research outcomes, methods, and applications, and aim to reveal central issues and integrate findings. We base our work on a representative source sample, combine conceptual and methodological formats to organize the review and present the results from a neutral perspective, addressing general scholars and public.

With respect to the history of the topic, web tracking was considered to be part of research on information seeking before the emergence of Web 2.0 (Taylor & Pentina, 2017). It was related to transaction-log analysis and can be dated back to the mid-1960s (Fourie & Bothma, 2007). In the literature review by Jansen & Pooch (2001), the focus was placed on web tracking for monitoring the use of databases, CD-ROM software and library catalogues.

This understanding of web tracking changed around the year 2006 (Fourie & Bothma, 2007). Since then, it refers to a set of techniques for websites to construct user profiles (Besson et al., 2014; Sanchez-Rola et al., 2016). Web tracking is nowadays also understood as a widespread Internet technique that collects user data for purposes of online advertisement, user authentication, content personalization, advanced

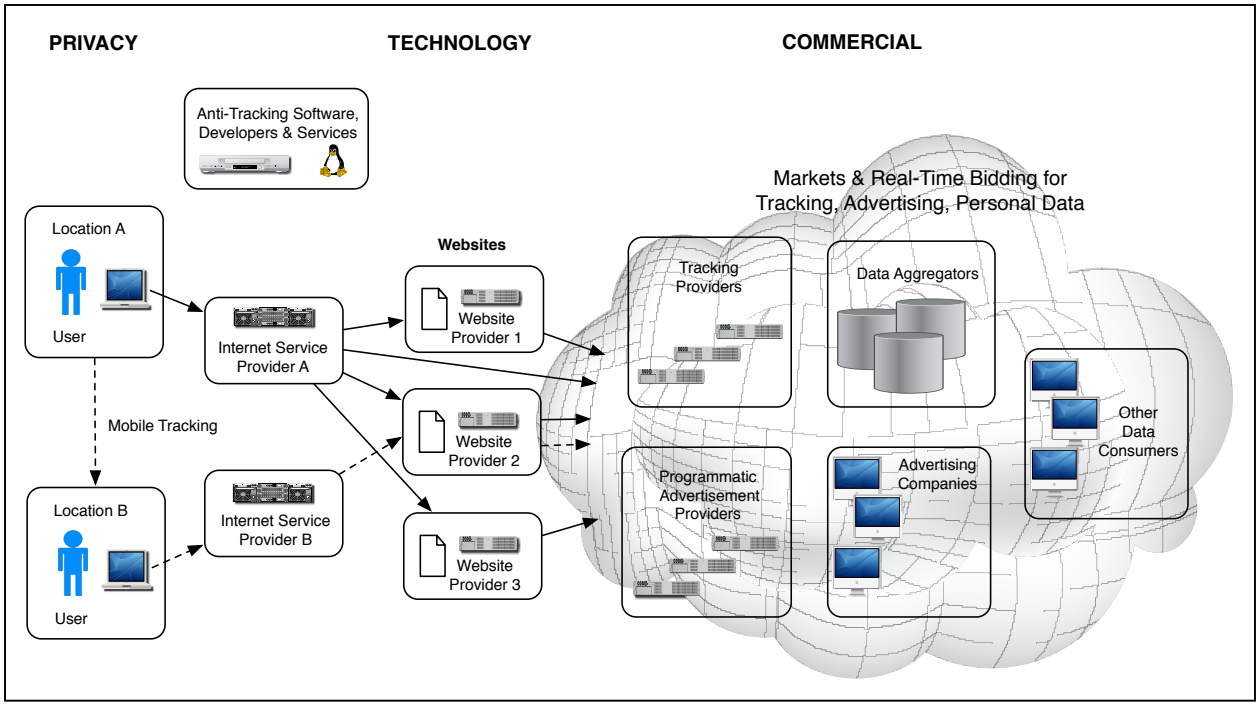

Figure 1: Overview and conceptualization of Web Tracking

2.2. Conceptualization of the topic

After the review scope has been defined, the research area is conceptualized to show what is known about the topic (Torraco, 2005). For this, we base on an informally collected set of starting literature gathered during earlier work, or recommended by literature repositories such as ResearchGate based on previous research interests. The main literature review, in contrast, focuses on documented and formal search strategies for verification and repeatability.

website analytics, social network integration, and website development (Sanchez-Rola et al., 2016; Mayer & Mitchell, 2012; Roesner et al., 2012; Fourie & Bothma, 2007). For these goals, web tracking allows third-party or first-party websites to keep track of users' browsing behavior, including browsing configuration and history (Sanchez-Rola et al., 2016).

A high-level overview and conceptualization of in web tracking, including the major stakeholders involved, is given in Figure 1. A user accesses websites from a local device through an Internet Service Provider (ISP). Websites and ISPs may include

tracking technology, either in-house or provided by third parties that provide tracking services for multiple sites (Pugliese, 2015), which enables cross-site tracking and data aggregation of individual browsing habits and interests. If the user switches to a different device or moves to another location, cross-device tracking (Brookman et al., 2017) and mobile tracking can be applied.

Tracking data is often used for targeted advertising (Roesner et al., 2012). This has created background markets for programmatic advertising, including real-time bidding for available advertising slots on the websites that are displayed to the user. Large-scale data aggregators and other data consumers are also interested to gather tracking and browsing data to enrich data profiles on individual web users. This creates major challenges for the protection of personal privacy. Anti-tracking software and services aim at reducing the privacy exposure to tracking mechanisms and infrastructure.

In summary, we identify three main aspects of web tracking research: technology, privacy, and commerce. In addition, we investigate what kind of research methodologies are applied in this field.

2.3. Literature search

The databases selected for literature acquisition included Google Scholar, EBSCOhost, IEEE Xplore, ScienceDirect, AIS Electronic Library (AISeL), Springer, ACM Digital Library. They were consulted in the title and keywords fields, except for Google Scholar and SpringerLink, where searches by keywords are not enabled. Table 1 shows the resulting number of hits without filtering restrictions, when working with "web tracking", "web security", "web privacy", "third party tracking", and "online advertising" as search items.

	Google Scholar	EBSCO host	Sprin–ger	ACM	Science Direct	AISel	IEEE Xpl.
Web tracking	1540000	10	128535	67326	123409	5238	2492
Web security	1710000	25	119553	80481	67899	7417	12778
Web privacy	2200000	5	44484	72155	23813	3948	2968
Third-party tracking	442000	1	82171	27325	61233	2298	155
Online Adverti–sing	1450000	427	38153	23518	56480	3798	800
Relevant by title	132	14	62	100	44	2	91
Relevant by keyword	-	2	-	159	22	1	15
Relevant by abstract	58	4	9	74	9	0	27
Total	**45**	**1**	**7**	**23**	**6**	**0**	**4**

Table 1. Literature by database

The articles were further checked for relevance based on their abstracts (see "Relevant by abstracts") and duplicates (see "Total"). Out of the sample of 86 articles, 58 could not be retrieved or were considered inappropriate after a thorough examination and, hence, were eliminated. As a result of backward / forward searches (Webster & Watson, 2002, Herz et al., 2010; vom Brocke et al., 2009), only three new articles were found. Finally, we obtained a total of 31 relevant articles for in-depth analysis.

2.4. Literature analysis and synthesis

In the next step, the collected literature was analyzed and synthesized. Firstly, the articles were investigated with respect to the research methodologies they apply. For these purposes, Wilde & Hess' (2007) consolidated spectrum of research methodologies in information systems (IS) was adopted (Table 2).

Secondly, the articles could be categorized as concentrating on technological, privacy, and commercial aspects (Table 3). Articles with focus on technological aspects mainly present how web tracking and anti-web tracking techniques work. The works concentrating on privacy aspects rather show the threats related to privacy. The papers oriented on commercial aspects survey the effectiveness of personalized advertising.

3. Results

3.1. Research methodologies

Research methodologies in the IS discipline can be generally distinguished in terms of the research paradigm into either behavioral science or design science (Wilde & Hess, 2007). The behavioral-science paradigm attempts to form and justify theories for explaining or predicting behavior of individuals or organizations (Hevner et al., 2004), whereas the design-science paradigm deals with developing and assessing IT artifacts (e.g., models, methods or systems) to enlarge their capabilities (Hevner et al., 2004; Wilde & Hess, 2007).

Table 2 shows that 18 of the retrieved articles are based on the design-science paradigm, while 13 follow the behavioral-science paradigm. The research methodologies within the design-science paradigm found in the retrieved articles include modeling (2 articles), prototyping (9), and argumentative-deductive analysis (7). The research methodologies within the behavioral-science paradigm involve grounded theory (5), qualitative-empirical cross-sectional analysis (5), and field study (3).

The identified prototypes are aimed to detect web tracking (Roesner et al., 2012) and to protect end users from these practices, e.g., *TrackMeOrNot* (Meng et al., 2016). Roesner et al. (2012) developed a client-side method for detecting five kinds of third-party trackers, classified based on how they manipulate browser state.

Researchers who used argumentative-deductive analysis mainly show pros and cons of web-tracking technology or the existing web tracking and anti-web tracking tools (e.g., Bujlow et al., 2015, 2017; Mayer & Mitchell, 2012; Pugliese, 2015).

Modeling was used to demonstrate web-tracking scenarios in practice. For instance, Puglisi et al. (2016) applied modeling to analyze how advertising networks build user footprints and how the suggested advertising reacts to changes in the user behavior.

Grounded theory was applied to observe how tracking and anti-tracking mechanisms work on the web, while qualitative-empirical cross-sectional analysis was based on interviews with focus on individuals' understanding and opinion on web tracking and personalized advertising. For instance, Melicher et al. (2016) collected browsing histories of 35 individuals and interviewed them about perceived benefits and risks of online tracking in the context of their own browsing behavior. In an example field study, Han et al. (2012) investigated how 20 participants were tracked over a time period of more than three weeks on their mobile phones.

Design Science	**Count**	**Publications**
Argumentative-deductive analysis	7	Bujlow et al. (2015, 2017); Clark et al. (2015); Cooper et al. (2013); Fourie & Bothma (2007); Jansen & Pooch (2001); Pugliese (2015); Sanchez-Rola et al. (2016)
Prototyping	9	Acar et al. (2014); Akkus et al. (2012); Besson et al. (2014); Englehardt & Narayanan (2016); Ikram et al. (2016); Meng et al., 2016; Roesner et al. (2012); Yamada et al. (2011); Stopczynski & Zugelder (2013)
Modeling	2	Gill et al. (2013); Puglisi et al. (2016)
Behavioral Science		
Grounded theory	5	Acar et al. (2013); Fourie & Bothma (2007); Javed (2013); Mayer & Mitchell (2012); Schelter & Kunegis (2016a);
Qualitative-empirical cross-sectional analysis	5	Agarwal et al. (2013); Budak et al. (2016); Melicher et al. (2015); Thode et al. (2015); Ur et al. (2012)
Field study	3	Falahrastegar et al. (2016); Han et al. (2012); Leung et al. (2016)

Table 2. Overview of research methodologies

3.2. Technological aspects

Bujlow et al. (2015, 2017) and Mayer & Mitchell (2012) provide a detailed overview of the existing web-tracking techniques. Bujlow et al. (2015, 2017) distinguish between five main groups of web-tracking techniques, which are based on sessions, client storage, client cache, fingerprinting, and other approaches. Mayer and Mitchell (2012) make a distinction between stateful and stateless web tracking techniques, depending on where data for user recognition is stored. With stateful tracking, the tracker stores the data required for user identification on the client side (Sanchez-Rola et al., 2016). With stateless tracking, the tracker collects users' browser and OS information to differentiate between them (Besson et al., 2014).

Stateful tracking techniques include cookies, *ETags*, and web storages (Pugliese, 2015). Cookies are used to store authentication data. There are many different types of them: Flash cookies are stored within the local storage used by Adobe Flash (Pugliese, 2015; Sanchez-Rola et al., 2016). Cookie syncing allows different trackers to share the same user identifiers (Sanchez-Rola et al., 2016). Supercookies and zombie cookies are stored on multiple storages and re-create themselves after being deleted.

Third-party cookies are used by domains which do not correspond to the currently visited website and are often caused by content provisioning of third parties (Pugliese, 2015). Also, they are the most common form of tracking (Sanchez-Rola, 2016). Web storages involve caches on the client device and can be accessed by browsers and plugins (Pugliese, 2015).

According to Mayer & Mitchell (2012), stateless tracking can be separated in active and passive fingerprinting. Fingerprinting is "the process of an observer or attacker uniquely identifying (with a sufficiently high probability) a device or application instance based on multiple information elements communicated to the observer or attacker" (Cooper et al., 2013, p. 7).

Aspects	Subject	Publications
Technology	*Web-tracking methods*	
	Stateful tracking	Acar et al. (2013, 2014); Besson et al. (2014); Bujlow et al. (2015, 2017); Englehardt & Narayanan (2016); Ikram et al. (2016); Mayer & Mitchell (2012); Pugliese (2015); Sanchez-Rola et al. (2016)
	Stateless tracking	Acar et al. (2014); Besson et al. (2014); Bujlow et al. (2015, 2017); Ikram et al. (2016); Mayer & Mitchell (2012); Pugliese (2015); Sanchez-Rola et al. (2016)
	Tracking behavior	
		Roesner et al. (2012)
	Web tracking on mobile devices	
		Han et al. (2015); Javed (2013); Leung et al. (2016); Pugliese (2015)
Privacy	*Problems*	
	Web-tracking methods increase	Clark et al. (2015); Falahrastegar et al. (2016); Ikram et al. (2016); Pugliese (2015); Sanchez-Rola et al. (2016)
	Ineffective tools	Acar et al. (2014); Melicher et al. (2016); Roesner et al. (2012); Stopczynski & Zugelder (2013); Yamada et al. (2011)
	Privacy invasion	Akkus et al. (2012); Besson et al. (2014); Bujlow et al. (2015, 2017); Clark et al. (2015); Cooper et al. (2013); Englehardt & Narayanan (2016); Gill et al. (2013); Leung et al. (2016); Mayer & Mitchell (2012); Melicher et al. (2015); Pugliese (2015); Sanchez-Rola et al. (2016); Yamada et al. (2011)
	Tools to protect privacy	
	Anti-web tracking tools	Akkus et al. (2012); Besson et al. (2014); Bujlow et al. (2015, 2017); Cooper et al. (2013); Englehardt & Narayanan (2016); Mayer & Mitchell (2012); Meng et al. (2016); Pugliese (2015); Roesner et al. (2012); Sanchez-Rola et al. (2016); Stopczynski & Zugelder, 2013
	Own developed anti-web tracking tools	Akkus et al. (2012); Besson et al. (2014); Englehardt & Narayanan (2016); Ikram et al. (2016); Meng et al. (2016); Roesner et al. (2015); Stopczynski & Zugelder (2013); Yamada et al. (2011)
Commercial	*Business*	
	Aspects	
	Business models for third-party tracking	Mayer & Mitchell (2012)
	Economical use of web tracking	Agarwal et al. (2013); Budak et al. (2016); Gill et al. (2013); Melicher et al. (2015); Puglisi et al. (2016); Thode et al. (2015); Ur et al. (2012)
	Web tracking as main income source	Clark et al. (2015); Fourie & Botchma (2007); Mayer & Mitchell (2012); Thode et al., (2015); Sanchez-Rola et al. (2016); Schelter & Kunegis (2016)
Other issues	*Do Not Track**	
		Acar et al. (2013); Agarwal et al. (2013); Akkus et al. (2012); Budak et al. (2016); Gill et al. (2013); Mayer & Mitchell (2012); Pugliese (2015); Roesner et al. (2012)

* Not included into literature review scope and therefore not further investigated.

Table 3. Results of literature analysis and synthesis

Browser fingerprinting is ideally suited to identify devices by using JavaScript (Pugliese, 2013). Canvas fingerprinting is used for device identification and uses the differences of pixel maps when rendering fonts and WebGL scenes in the browser (Pugliese, 2015; Sanchez-Rola et al., 2016). Pugliese (2015) also mentions behavioral biometric features, namely those dynamics that occur when typing, moving and clicking the mouse, or touching a touch screen. Such behavioral biometric features can be used to improve stateless tracking.

Tracking methods make it difficult to block all third-party content. Furthermore, it is necessary to accept some third-party content to ensure web site functionality (Stopczynski & Zugelder, 2013). Falahrastegar et al. (2016) found that users are even being tracked regardless of their profile condition (logged-in or logged-out).

Increasing awareness of users on data protection and privacy led to browser settings and extensions to delete or prevent certain kinds of cookies and trackers, but new methods are constantly being developed and changed continuously in order to track and identify users (Falahrastegar et al., 2016). An example of this trend is the emergence of various user-tracking mechanisms.

According to Roesner et al. (2012), there are five tracking behavior types (Category A-E in Table 4). In category A, entitled 'analytics', the third-party tracker tracks users only within one web site (e.g., Google Analytics).

Cat.	Name	Profile Scope	Summary	Example	Visit directly
A	Analytics	Within Site	Serves as third-party analytics engine for sites	Google Analytics	No
B	Vanilla	Cross-Site	Uses third-party storage to track users across sites	Double Click	No
C	Forced	Cross-Site	Forces user to visit directly (e.g., via popup or redirect)	Insight Express	Yes, Forced
D	Referred	Cross-Site	Relies on a B, C, or E tracker to leak unique identifiers	Invite Media	No
E	Personal	Cross-Site	Visited directly by the user in other contexts	Facebook	Yes

Table 4. Classification of tracking behavior (Roesner et al., 2012)

In category B, or 'vanilla', the third-party tracker relies on available third-party storages to track users across web sites. In category C, 'forced', the cross-site tracker makes users visit its web site domain directly (e.g., via popup, redirect), turning into a first-party position. Within category D, called 'referred', the tracker reveals unique identifiers from B, C or E trackers to track users across sites, instead of storing them on its own. In category E, known as 'personal', the cross-site tracker is called directly in other contexts (e.g., Facebook). Within this framework, only categories B and E are mutually exclusive, whereas other categories can be shared.

The literature analysis has also shown that end users are not only tracked on the web, but also on their mobile phones (Han et al., 2012; Leung et al., 2016). According to the findings of Han et al. (2012), with 20 Android smartphone users observed over a time period of more than three weeks, tracking took place on every third visited website. Leung et al. (2016) surveyed the differences of web and mobile tracking and argue that there is a larger privacy threat on mobile phones due to additional privacy-critical information, e.g., end-users' locations, their phone number and contacts, call and email histories, and more. Compared to tracking on the web, common practices on mobile devices are largely unknown and not well understood (Han et al., 2012; Leung et al., 2016), except for initial studies (Eubank et al. 2013).

According to Acar et al. (2014), websites should consider integrating user protection more deeply into the browser. Clark et al. (2015) suggest disrupting the linkability in tracker databases. Sanchez-Rola et al. (2016) suggest that spoofing a user profile could prevent web tracking. However, this approach could be counterproductive since even those attempts of hiding one's identity can also be used for fingerprinting. Ikram et al. (2016) underline that many tracking tools are based on JavaScript; therefore, it would be useful to develop a corresponding filtering mechanism.

3.2. Privacy aspects

Several of the identified articles discuss privacy invasion caused by the use of web tracking (Table 3). According to Mayer & Mitchell (2012), a web-browsing history is inextricably linked to personal information. The websites a user visits can reveal her location, interests, purchases, employment status, sexual orientation, financial challenges, medical conditions, news consumption, and can be used as instrument for mass surveillance by intelligence agencies, and more (Mayer & Mitchell, 2012; Schelter & Kunegis, 2016a). Furthermore, Mayer & Mitchell (2012) mention ascertained information by web tracking that is very personal, e.g., menopause, getting pregnant, repairing bad credit, debt relief or how often a user drinks, smokes and consumes drugs.

These examples show that web tracking can have considerably negative consequences for end users. Schelter & Kunegis (2016a) discovered that even though the rate of third-party tracked websites among those with highly privacy-critical content is lower than for other websites (60% versus 90%), the majority of such websites does contain trackers.

There exist several techniques to protect the privacy of the user, such as third-party cookie blocking, clearing the client-side state, blocking popups, *AdBlock Plus*, *Adblock Edge*, *Ghostery*, *BetterPrivacy*, *Site Isolation, EFF's Privacy Badger* and private browsing mode (Ikram et al., 2016). Bujlow et al. (2015, 2017) provide an overview of anti-web tracking tools for a specific web tracking technique. Such tools are available for stateful and stateless web tracking, but do not block web tracking effectively and are complicated to use for end users (Acar et al., 2014). For instance, a private browsing mode prevents specific data from falling into the hands

of other users on the same computer; it also prevents long-term tracking based on stateful techniques. But as long as JavaScript is enabled or certain plugins are installed, device fingerprinting cannot be prevented (Pugliese, 2015). Therefore, Sanchez-Rola et al. (2016) suggest disabling other secondary features used in web tracking. This would constitute a more promising approach because the number of websites that rely on their functionality is smaller. Disabling third-party cookies can be effective (Pugliese, 2015). *Adblock Edge* suppresses the display of advertising websites. *BetterPrivacy* deletes supercookies and thereby prevents long-term tracking. *Ghostery* blocks various types of (third-party) cookies and trackers (Pugliese, 2015). Sanchez-Rola et al. (2016) and Melicher et al. (2016) further identify completely functional anti-tracking web browsers, e.g., *FlowFox* (De Groef et al., 2012), *TrackingFree* and *Privaricator* (Nikiforakis et al., 2015). However, the main problem of these methods is that they only take into account certain fields and privacy attacks due to the computational complexity of tracking (Sanchez-Rola et al., 2016). Even if combined with further communication anonymizers, issues of usability remain (Brecht et al, 2011). Akkus et al. (2012) suggest web analytics without tracking. Interestingly, even those prototypes that were designed to protect privacy cannot protect end users against all types of web tracking.

Moreover, willingness to adopt privacy-enhancing tools can be dependent on user personality traits (Brecht et al, 2012). Thode et al. (2015) interviewed 20 German participants without technical skills and found that participants are frightened and tend to avoid using the Internet completely after being informed about how often they are tracked. This raises concerns about the lack of privacy protection and can influence the economy of web tracking (Thode et al., 2015). Ur et al. (2012) also underline the necessity for more privacy protection or at least more transparency. Agarwal al. (2013) and Thode et al. (2015) discuss the failure of existing methods within the advertising industry to raise awareness, knowledge, and trust on third-party tracking. The latter suggest charging an independent, non-commercial organization as a widely known and trusted third party to certify online tracking methods.

Sanchez-Rola et al (2016) list tools (*Adnostic*, *PrivAd*, *RePriv* and *OblivAd*) that are proposed for analytics and targeting, preserving users' privacy in the context of online behavioral advertisement.

3.3. Commercial aspects

Only few details on the commercial aspects of web tracking could be identified in the literature, although online advertising was used as a keyword for article selection. In line with this finding, Gill et al. (2013, p. 1) argues that "little is known about the economics of online advertising, chiefly the economics of collecting and using personal information about users for facilitating targeted advertising".

Generally, online companies use web tracking for website optimization, e.g., with regards to usability and user browsing experience (Melicher et al., 2016; Pugliese, 2015; Sanchez-Rola et al., 2016). Advertising companies use web tracking mechanisms primarily to show personalized, tailored advertisements to their users (Clark et al., 2015). Mayer & Mitchell (2012) speak about six common high-level business models related to third-party websites: advertising companies, analytics companies, analytics services, social networks, content providers, frontend services, and hosting platforms. For advertising companies, there are three main models: direct buy, ad networks, and ad exchanges. Direct buy is the oldest model of online advertising and remains the dominant model for search-engine and social-network advertising. Ad networks are the largest and most widely used intermediaries in online advertising. Here, advertisers and first-party websites do not deal directly, and advertisers can easily place ads with many publishers. With ad exchanges, bids are made via many advertising networks. These ad exchanges led to a number of intermediary business models that exist in the current exchange ecosystem (e.g., demand-side platforms, supply-side platforms, data providers). Out of those, data providers are the most interesting for this research because they sell ad-targeting data to advertisers in real time.

The second business model for third-party websites involves analytics services, which provide tools for websites to better understand their visitors, including demographics, user agents, and content views and interactions. Examples for such services include *Adobe Analytics, Quantcast* and *Google Analytics*.

The third business model for third-party websites is social networks or social integration. Here, "social integration enables websites to offer personalized content and single sign-on to social network users" (Mayer & Mitchell, 2012, p. 419). According to Mayer & Mitchell (2012), social integration is practiced the most on first-party social networks. Most prominent examples are the Facebook like button, Twitter tweet and the Google +1 button. The social networks offer their services for free to increase user engagement and conduct market research.

Social integration has led to several intermediary business models, e.g., social sharing aggregation with services such as *AddThis*, *ShareThis*, and *Meebo*. Here, widgets are offered for free to websites that enable

users to share content with dozens of social networks and generate revenue by collecting and selling tracking and usage data for ad targeting and market research.

The fourth model, content providers, involves the hosting of videos, maps, news, weather, stocks, and other media for embedding into websites. Examples for this include YouTube and Associated Press and also Google, Facebook and Amazon (Bujlow et al., 2015, 2017). The fifth model, frontend services, "host JavaScript libraries and APIs that speed webpage loads (e.g., Google Libraries API) and enable new page functionality (e.g., Google Feed API)" (Mayer & Mitchell, 2012).

Within the last model, hosting platforms maintain services that support publishers in spreading their own content, e.g., blog platforms or content distribution networks. In practice, many services cut across business models, and novel business models are evolving.

Advertising that relies on web tracking techniques is often called *online behavioral advertising* (OBA) Here, advertising networks profile a user based on her online activities in terms of the websites she visits over time (Ur et al., 2012). A user's browsing history is retrieved based on her identifier on the visited websites within the advertisement network (Sanchez-Rola et al., 2016). Advertising networks use this history to show ads that are more likely to be of interest to a particular user (Ur et al., 2012).

Fingerprinting (stateless tracking) has become an increasingly common practice used by advertisement enterprises (Sanchez-Rola et al., 2016). For companies that use web tracking, "efficient and successful advertising relies on predicting users' actions and tastes to a range of products to buy" (Puglisi et al., 2016). Interestingly, the existing tracking tools –both stateful and stateless – fail to address the complexity of buying decisions and, therefore, perform poorly at supporting desired behavior predictions (Melicher et al., 2016).

Gill et al. (2013) find in their study that better privacy tools, namely to block third parties, would decrease overall revenue by 75%. However, Ur et al. (2012) claim online consumers feel less discomforted with personalized advertising when being properly informed about the usage of non-personally identifiable information for OBA. This appears to be in line with Thodes et al.'s (2015) suggestion to make web tracking more transparent for end users.

Unfortunately, the identified research articles do not report the exact techniques that are used for OBA. Ur et al. (2015) mention that tracking can be exercised in multiple ways and point out that the tracking methods for OBA maintain a unique identifier on a user's computer over time.

3.4. Research agenda

Despite a well-established systematic literature review framework applied in this work, some relevant articles on web tracking might have not been part of the present analysis, among other reasons possibly due to the keyword selection (Herz et al., 2010).

The state-of-the-art research on web tracking was analyzed and presented with regards to technological, privacy and commercial aspects. Other perspectives were left for consideration within future research projects, e.g., discussions on global variations of web tracking (Mayer & Mitchell, 2012), customers' perceptions(Agarwal et al., 2013; Melicher et al., 2015; Thode et al., 2015; Ur et al., 2012) or compromises with them (Mayer & Mitchell, 2012).

The analysis within the framework by Wilde & Hess (2006) shows that there is no research based on simulation, action research, formal-deductive analysis and conceptual-deductive analysis within the design science research paradigm, as well as on case study, labor study, quantitative-empirical cross-sectional analysis and ethnographic analysis within the behavioral science research paradigm.

This research could generally confirm previously reported results on web tracking (e.g., Roesner et al., 2012; Falahrastegar et al., 2016) calling for more clarity about the working principles of web tracking, assessments of its prevalence, and the scope of constructible user browsing profiles.

Insights into the commercial side of web tracking were found to be rather limited. Future research avenues could show which techniques are used for OBA and how personalized offers can be improved by using tracking techniques. Behavioral biometric features are rather mentioned in terms of their capabilities to improve tracking and, hence, provide a field for more detailed explorations.

There is so far only a small number of research articles on mobile web tracking (Han et al., 2012; Leung et al., 2016), which could adopt physical tracking of users through apps on their GPS-enabled smartphones, integrating location data to the already rich set of user information interests, which is of particular interest for displaying advertisements based on precise user location. Since such tracking with respect to the physical world can create severe privacy impacts, research on location privacy from field such as pervasive computing (Beresford & Stajano, 2003) should be integrated with mobile (web) tracking research.

From a privacy perspective, more efficient end user protection against all forms of tracking would be required. The literature shows an arms race between novel tracking and fingerprinting technology and

privacy defense. Moreover, most of the privacy-enhancing mechanisms have not yet been rigorously tested, broadly implemented, or adopted by users. Reconciliation of commercial interests with privacy, and the limits, constitutes another important avenue for future research.

4. Conclusion

We have conducted a structured literature review based upon an established methodological framework, and provided an overview of the applied research methods. We further evaluated the articles with regards to technological, privacy and commercial aspects. Our research shows that future research efforts could focus on mobile web tracking, protecting end-users effectively against web tracking, and how commercial and privacy interests could be reconciled.

5. References

[1] G. Acar, C. Eubank, S. Englehardt, M. Juaerez, A. Narayanan, and C. Diaz, "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild", 21st CCS (2014), Scottsdale, USA, ACM, pp. 674-689.

[2] G. Acar, M. Juarez, N. Nikiforakis, C. Diaz, S. Gürses, F. Piessens, and B. Preneel, "FPDetective: Dusting the Web for Fingerprinters", 20th CCS (2013), Berlin, Germany, ACM, pp. 1129-1140.

[3] L. Agarwal, N. Shrivastava, S. Jaiswal, and S. Panjwani, "Do Not Embarrass: Re-Examining User Concerns for Online Tracking and Advertising", 9th Symp. on Usable Privacy and Security (2013), Newcastle, UK, pp. 8:1-8:13.

[4] I.E. Akkus, R. Chen, M. Hard, P. Francis, and J. Gehrke, "Non-tracking Web Analytics", 19th CCS (2012), Raleigh, USA, ACM.

[5] M.J. Baker, "Writing a Literature Review", The Marketing Review (2000), 1, 2, pp. 219-247.

[6] B. Bender, B. Fabian, S. Lessmann, and J. Haupt, "E-Mail Tracking: Status Quo and Novel Countermeasures", ICIS (2016), Dublin, Ireland, AIS.

[7] A.R. Beresford and F. Stajano: Location Privacy in Pervasive Computing. IEEE Pervasive Computing (2003), pp. 46-55

[8] F. Besson, N. Bielova, and T. Jensen, "Hybrid Information Flow Monitoring Against Web Tracking", 26th Computer Security Foundations Symp. (2014), New Orleans, USA, IEEE Computer Society, pp. 240-254.

[9] F. Brecht, B. Fabian, S. Kunz, and S. Müller, "Are You Willing to Wait Longer for Internet Privacy?", ECIS (2011), Helsinki, Finland.

[10] F. Brecht, B. Fabian, S. Kunz, and S. Müller, "Communication Anonymizers: Personality, Internet Privacy Literacy and Their Influence on Technology Acceptance", ECIS (2012), Barcelona, Spain.

[11] J. Brookman, P. Rouge, A. Alva, and C. Yeung, "Cross-Device Tracking: Measurement and Disclosures," Proc. Privacy Enhancing Technologies (2017), 2, pp. 133–148.

[12] C. Budak, S. Goel, J. Rao, and G. Zervas, "Understanding Emerging Threats to Online Advertising", Conf. on Economics and Computation (2016), Maastricht, Netherlands, ACM, pp. 561-578.

[13] T. Bujlow, V. Carela-Espanol, J. Sole-Pareta, and P. Barlet-Ros, "Web Tracking: Mechanisms, Implications, and Defenses", arXiv:1507.07872v1 (2015).

[14] T. Bujlow, V. Carela-Español, J. Solé-Pareta, and P. Barlet-Ros, "A Survey on Web Tracking: Mechanisms, Implications, and Defenses", Proceedings of the IEEE (2017), 105, 8, pp. 1476-1510.

[15] S. Clark, M. Blaze, and J. Smith, "Smearing Fingerprints: Changing the Game of Web Tracking with Composite Privacy", Security Protocols XXIII – 23rd Int. Workshop (2015), Cambridge, UK, Springer, pp. 178-182.

[16] A. Cooper, H. Tschofenig, J. Peterson, J. Morris, M. Hansen, and R. Smith, "Privacy Considerations for Internet Protocols", https://tools.ietf.org/html/rfc6973 (2013).

[17] H.M. Cooper, "Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews", Knowledge in Society (1988), 1, 1, pp. 104-126.

[18] S. Englehardt, and A. Narayanan, "Online Tracking: A 1-million-site Measurement and Analysis", 23rd CCS, Vienna, Austria (2016), ACM, pp. 1388-1401.

[19] T. Ermakova, A. Hohensee, I. Orlamünde, and B. Fabian, "Privacy-Invading Mechanisms in E-Commerce – A Case Study on German Tourism Websites", International Journal of Networking and Virtual Organisations (2017).

[20] C. Eubank, M. Melara, D. Perez-Botero, A. Narayanan: "Shining the Floodlights on Mobile Web Tracking: A Privacy Survey", Web 2.0 Security & Privacy (2013), San Francisco, USA.

[21] B. Fabian, B. Bender, and L. Weimann, "E-Mail Tracking in Online Marketing: Methods, Detection, and Usage", 12th Int. Conf. Wirtschaftsinformatik (2015), Osnabrück, Germany.

[22] M. Falahrastegar, H. Haddadi, S. Uhlig, and R. Mortier, "Tracking Personal Identifiers across the Web, Int. Conf. on Passive and Active Network Measurement (2016), Heraklion, Greece, LNCS, Springer, pp. 30-41.

[23] J. Fernandez-Tapia, "Rigorous Budget Allocation for Programmatic Ad-Buying", Technical Report (2016), https://www.researchgate.net/publication/286649541.

[24] I. Fourie, and T. Bothma, "Information Seeking: An Overview of Web Tracking and the Criteria for Tracking Software", Aslib Proceedings (2007), 59, 3, pp. 264 - 284.

[25] P. Gill, V. Erramilli, A. Chaintreau, B. Krishnamurthy, and D. Papagiannaki, "Follow the Money: Understanding Economics of Online Aggregation and Advertising", Privacy Enhancing Technologies (2013), pp. 141-148.

[26] R. Gomer, E. M. Rodrigues, N. Milic-Frayling, and M. C. Schraefel, "Network Analysis of Third Party Tracking: User Exposure to Tracking Cookies through Search", IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (2013).

[27] W. De Groef, D. Devriese, N. Nikiforakis, F. Piessens, "FlowFox: A Web Browser with Flexible and Precise Information Flow Control", 19th CCS, Raleigh, USA, (2012), ACM.

[28] A. Hamed, H. K.-B. Ayed, M. A. Kaafar, and A. Kharraz, “Evaluation of Third Party Tracking on the Web”, Int. Conf. for Int. Techn. and Sec. Transact. (ICITST) (2013).
[29] S. Han, J. Jung, and D. Wetherall, “A Study of Third-Party Tracking by Mobile Apps in the Wild”, Technical Report (2012), University of Washington.
[30] A.R. Hevner, S.T. March, J. Park, and S. Ram, “Design Science in Information Systems Research”, MIS Quarterly 28 (2004), 1, pp. 75-105.
[31] T.P. Herz, F., Hamel, F. Uebernickel, and W. Brenner, “Deriving a Research Agenda for the Management of Multisourcing Relationships Based on a Literature Review”, 16th AMCIS, Lima, Peru (2010), AIS, pp. 1-11.
[32] M. Ikram, H.J. Asghar, M.A., Kaafar, A. Mahanti, and B. Krishnamurthy, “Towards Seamless Tracking-Free Web: Improved Detection of Trackers via One-class Learning”, Privacy Enhancing Technologies (2016), 2017, 1, pp. 79-99.
[33] B.J. Jansen, and U. Pooch, “A Review of Web Searching Studies and a Framework for Future Research”, Journal of the ASIS&T (2001), 52, 3, pp. 235-246.
[34] A. Javed, “POSTER: A Footprint of Third-Party Tracking on Mobile Web”, 20th CCS, Berlin, Germany, (2013), ACM, pp. 1441-1444.
[35] T. Libert, “Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites”, International Journal of Communication (2015), 9, pp. 3544-3561.
[36] C. Leung, J. Ren, D. Choffnes, and C. Wilson, “Should You Use the App for That? Comparing the Privacy Implications of App- and Web-based Online Service”, Internet Measurement Conference (2016), Santa Monica, USA, ACM, pp. 365-372.
[37] J.R. Mayer, and J.C. Mitchell, “Third-Party Web Tracking: Policy and Technology”, IEEE Symposium on Security and Privacy (2012), San Francisco, USA, IEEE, pp. 413-427.
[38] W. Melicher, M. Sharif, J. Tan, L. Bauer, M. Christodorescu, and P.G. Leon, “(Do Not) Track Me Sometimes: Users’ Contextual Preferences for Web Tracking”, Privacy Enhancing Technologies (2015), 2016, 2, pp. 135–154.
[39] W. Meng, B. Lee, X. Xing, and W. Lee, “TrackMeOrNot: Enabling Flexible Control on Web Tracking”, 25th WWW (2016), Montréal, Canada, ACM, pp. 99-109.
[40] I.D. Mitchell, “Third-Party Tracking Cookies and Data Privacy”, SSRN Electronic Journal (2012).
[41] A. Narayanan, and V. Shmatikov, “De-anonymizing Social Networks”, IEEE Symp. Security and Privacy (2009).
[42] N. Nikiforakis, W. Joosen, and B. Livshits, “Privaricator: Deceiving Fingerprinters with Little White Lies”, 24th WWW (2015), Florence, Italy, ACM, pp. 820–830.
[43] J. O'Connell, “The State of Programmatic Media”, https://research.adexchanger.com/the-state-of-programmatic-media.html (2014).
[44] J. Parra-Arnau, “Pay-per-Tracking: A Collaborative Masking Model for Web Browsing”, Information Sciences (2017), 385–386, pp. 96–124.
[45] S. E. Peacock, “Prized Assets for Web Tracking”, International Conference on Social Media & Society (2015).
[46] G. Pugliese, “Web Tracking: Overview and applicability in digital investigations”, Information Technology (2015), 57, 6, pp. 366–375.
[47] S. Puglisi, D. Rebollo-Monedero, and J. Forne, “On Web User Tracking: How Third-Party HTTP Requests Track Users’ Browsing Patterns for Personalised Advertising”, 15th IFIP MEDHOCNET (2016), IEEE, pp. 1-6.
[48] F. Roesner, T. Kohna, and D. Wetherall, “Detecting and Defending Against Third-Party Tracking on the Web”, 10th Int. Conf. on Web and Social Media (2012), San Jose, USA, pp. 155-168.
[49] I. Sanchez-Rola, X. Ugarte-Pedrerp, I. Santos, and P.G. Bringas, “The Web is Watching You: A Comprehensive Review of Web-Tracking Techniques and Countermeasures”, Logic Journal of the IGPL (2016), 25, 1, pp. 18-29.
[50] S. Schelter, and J. Kunegis, “Tracking the Trackers: A Large-Scale Analysis of Embedded Web Trackers”, 10th Int. Conf. on Web and Social Media (2016), Cologne, Germany, AAAI Press, pp. 679-682.
[51] S. Schelter, and J. Kunegis, “On the Ubiquity of Web Tracking: Insights from a Billion-Page Web Crawl”, arXiv:1607.07403 (2016b).
[52] M. Stange, and B. Funk, “Real-Time Advertising”, Business & Information Systems Engineering (2014), . 6, . 5, pp. 305-308.
[53] M. Stopczynski, and M. Zugelder, “Reducing User Tracking through Automatic Web Site State Isolation”, 17th Int. Conf. on Information Security (2013), Hong Kong, China, pp. 309-327.
[54] D. Taylor, and I. Pentina, “Guest Editorial: Branding in the Era of Web 2.0 (and beyond)”, Journal of Product & Brand Management (2017), 26, 4, pp. 341-341.
[55] W. Thode, J. Griesbaum, and T. Mandl, “’I would have never allowed it’: User Perception of Third-party Tracking and Implications for Display Advertising”, 14th Int. Symp. on Information Science (2015), Zadar, Croatia, pp. 445-456.
[56] R.J. Torraco, “Writing Integrative Literature Reviews: Guidelines and Examples”, Human Resource Development Review (2005), 4, 3, pp. 356-367.
[57] B. Ur, P.G. Leon, L.F. Cranor, R. Shay, and Y. Wang, “Smart, Useful, Scary, Creepy: Perceptions of online behavioral advertising”, 8th Symp. on Usable Privacy and Security (2012), Washington, USA, ACM, Article No. 4.
[58] J. vom Brocke, A. Simons, B. Niehaves, K. Riemer, R. Plattfaut, and A. Cleven, “Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process”, ECIS (2009), Verona, Italy, pp. 2206-2217.
[59] J. Webster, and R.T. Watson, “Analyzing the Past and Prepare for the Future: Writing a Literature Review”, MIS Quarterly (2002), 26, 2, pp. xiii-xxiii.
[60] T. Wilde, and T. Hess, “Research Methods in ‚Wirtschaftsinformatik’ – An Empirical Study“, Wirtschaftsinformatik (2007), 49, 4, pp. 280–287.
[61] A. Yamada, M. Hara, and Y. Miyake, “Web Tracking Site Detection Based on Temporal Link Analysis and Automatic Blacklist Generation”, IPSJ Journal (2011), 52, 2, pp. 633-644.