Web Tracking Technologies and Protection Mechanisms

Nataliia Bielova
Université Côte d'Azur, Inria
Sophia Antipolis, France
nataliia.bielova@inria.fr

ABSTRACT

Billions of users browse the Web on a daily basis, leaving their digital traces on millions of websites. Every such visit, every mouse move or button click may trigger a wide variety of hidden data exchanges across multiple tracking companies. As a result, these companies collect a vast amount of user's data, preferences and habits, that are extremely useful for online advertisers and profitable for data brokers, however very worrisome for the privacy of the users.

In this *3-hours tutorial* we will cover the vide variety of Web tracking technologies, ranging from simple cookies to advanced cross-device fingerprinting. We will describe the main mechanisms behind web tracking and what users can do to protect themselves. Moreover, we will discuss solutions Web developers can use to automatically eliminate tracking from the third-party content they include in their applications. This tutorial will be of interest to a *general audience* of computer scientists, and *we do not require any specific prerequisite knowledge* for attendees.

We will cover the following tracking mechanisms:

- third-party cookie tracking, and other stateful tracking techniques that enables tracking across multiple websites [26, 28, 29],
- cookie respawning that is used to re-create deleted user cookies [7, 27],
- cookie synching that allows trackers and ad agencies to synchronise user IDs across different companies [13, 24],
- browser fingerprinting, including Canvas, WebRTC and AudioContext fingerprinting [6, 11, 13, 21]
- cross-browser device fingerprinting, allowing trackers to recognise users across several devices [10].

We will then demonstrate prevalence of such techniques on the Web, based on previous research [6, 13, 17, 22, 26, 31]. We will present the advertisement ecosystem and explain how Web technologies are used in advertisement, in particular in Real-Time-Bidding (RTB). We will explain how cookie synching is used in RTB and present recent analysis on how much a user's tracking data is worth [24]. We will discuss the mechanisms the website owners use to automatically interact with the ad agencies [23], and explain its consequences on user's security and privacy.

To help users protect themselves from Web tracking, we will give an overview of existing solutions. We'll start with the browser settings, and show that basic third-party cookie tracking is still possible even in the private browser mode of most common Web browsers. We then present privacy-protecting browser extensions and compare how efficient they are in protection from Web tracking [19]. Then, we'll present possible protection mechanisms based on browser randomisation [15, 16] to protect from advanced fingerprinting techniques.

Finally, we will present solutions for Web developers, who want to include third-party content in their websites, but would like to automatically remove any tracking of their users. In particular, we will discuss simple solutions that exist today for social plugins integration [25], and propose more advanced server-side based solutions that are a result of our own research [30].

KEYWORDS

web tracking; surveillance; online privacy; big data

BIO

Nataliia Bielova is a Research Scientist at Inria, French National Institute for Research in Computer Science and Automation. Nataliia is internationally known for her work on applying formal methods to security and privacy of web browsers. Her main interest is privacy- and transparency-enhancing technologies for Web applications. She works on measurement, detection and prevention of web tracking, including advanced behaviour-based fingerprinting [14]. Nataliia received the French Doctoral Supervision and Research Award (PEDR) in 2017. Before obtaining her permanent position in 2013, Nataliia was a postdoctoral researcher at Inria Rennes from 2012 to 2013, where she worked on automatic detection of web tracking scripts using program analysis [8]. She received her PhD in Computer Science from the University of Trento, Italy in 2011.

1 HOW DOES WEB TRACKING WORK?

Web tracking is the practice when some content ("trackers") embedded in a webpage recognises the users visiting the page. Differently from analytics, trackers are usually

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
CCS '17, October 30-November 3, 2017, Dallas, TX, USA
© 2017 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-4946-8/17/10.
https://doi.org/10.1145/3133956.3136067

[1] This work has been partially supported by the ANR project AJACS ANR-14-CE28-0008.

FOLEON EX1018
IPR2026-00428

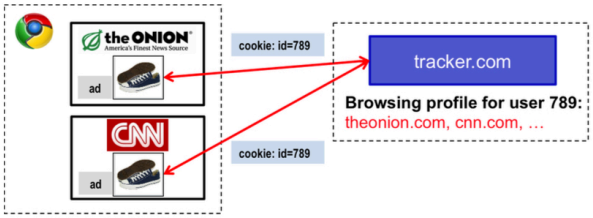

Figure 1: Basic cookie-based tracking from [17]. The third-party domain `tracker.com` recognises a user with a cookie "789" on sites that embed content from `tracker.com`.

"third-party" meaning that they belong to a different domain than the hosting webpage. Because trackers are originating from a third-party domain, they are able recognise user's across different websites where these trackers are embedded [17, 18, 26].

The basic tracking mechanism based on cookies is shown in Figure 1. When a user visits `theonion.com`, the browser loads additional third-party content from `tracker.com`. This content sets up a browser cookie with value "789" that is automatically stored in the browser. Upon a later request to a different website `cnn.com`, the third-party tracker `tracker.com` will recognise the same user "789" and thus will learn that this user has been to `theonion.com` and `cnn.com`.

Web tracking technologies are largely divided into two broad categories: *stateful* and *stateless*. Stateful tracking techniques store information on the user's computer and later retrieve it to recognise the user. Third-party cookies are the most prevalent online tracking technique, and in recent years researchers found out that cookies can be "respawned" even if the user deleted them [7, 27]. Moreover, cookies get synchronised among different data brokers in order to map and exchange user's profiles [13, 24]. Several groups of researchers have reported on the usage of different stateful trackers on popular websites [7, 17, 26, 28].

Stateless technologies allow trackers to recognise users without storing any information on the user's machine. Device fingerprinting collects information about the user's browser and OS properties, and can distinguish users by these characteristics [11]. In 2010, Eckersley first demonstrated that the technology is effective (see Panopticlick project [12]). Researchers later on have discovered that device fingerprinting started being used by tracking companies [6, 21]. Moreover, because of the advance of web standards and HTML5, new more advanced techniques became possible, that use Canvas API, WebRTC and AudioContext fingerprinting [13]. Recently, more fingerprinting techniques were discovered, that allow to track users across devices [9, 10]. In another recent project, researchers demonstrate that web browsers may be detected via the extensions that the user installs and websites where the user is logged in [14].

2 ADVERTISEMENT AND REAL-TIME-BIDDING

Real-Time-Bidding (RTB) is one of the main mechanisms that advertisement agencies are using to target users online. The advertisement ecosystem has substantially grown in the last 10 years, building on standard and more advanced web tracking technologies. We will show how fast this ecosystem has grown over the last years, and present main Web technologies used in advertisement. Researchers have detected that cookie synching is the main component of RTB and analysed how much a user's tracking data is worth in RTB: it is often sold for less than $0.0005 [24]. It was revealed that in RTB, the website owners and the ad agencies often tightly cooperate [23]. Such collaboration sometimes leads to bypassing basic Web security protections implemented by the browser, such as the Same-Origin-Policy. We discuss potential security issues raised by such collaboration.

3 PROTECTION FROM WEB TRACKING

The most common way to prevent most of the cookie tracking is to use the browser configuration and explicitly block third-party cookies. Even though research shows that in presence of other tracking techniques, cookies can be "respawned", the fact that the browser itself does not set, and moreover, does not send any third-party cookies, is already the first step to protect yourself. Private browsing modes in modern browser provide some protection as well, however they do not disable third-party cookies.

Another line of protection is by browser extensions. *AdBlock Plus* [1], *Ghostery* [3], *uBlock* [5], *Disconnect* [2] and *Privacy Badger* [4] are the most used privacy extensions in 2016 [19]. AdBlock Plus and uBlock rely on two community-driven rulesets that define whether a certain third-party content is a tracker. However, they do not provide full protection because if a tracker sets up a new domain not present in these rulesets, tracking via such domain will not be blocked. Ghostery and Disconnect create blocking rules internally. Differently from rule-based approaches, Privacy Badger uses heuristics to automatically detect third-party trackers on-the-fly. Even though none of these extensions can guarantee a 100% protection, they however disallow well-known companies to track you on the Web. Recently, researchers have measured the effectiveness of all these extensions on the top 200,000 websites [19].

Most of the solutions presented above do protect from a significant number of known trackers, but may be less effective against device fingerprinting. To protect yourself from this advanced stateless tracking, the simplest way is to disable JavaScript, however this may be not very practical. Several web browsers, proposed by researchers, aim at protecting from fingerprinting through the randomisation of browser properties [15, 16, 20].

Finally, website developers might also be interested in protecting their users from web tracking. This choice may be important either due to ethical reasons, or requirement

FOLEON EX1018
IPR2026-00428

for legal compliance, such as with upcoming EU ePrivacy regulation, which may make website owners liable for the third-party tracking present on their websites. *Social share privacy* platform [25] allows website developers to include social widgets, but disable any request to a third party until the used clicks on the button. Researchers have recently proposed *a server-side technique to protect against web tracking*: it's based on setting up two additional servers that rewrite and redirect the original web application requests so that third-party tracking is automatically removed from the third-party content [30].

REFERENCES

[1] 2017. Adblock Plus – Surf the web without annoying ads! (2017). https://adblockplus.org/.

[2] 2017. Disconnect. (2017). https://disconnect.me/.

[3] 2017. Ghostery. (2017). https://www.ghostery.com/.

[4] 2017. Privacy Badger - Electronic Frontier Foundation. (2017). https://www.eff.org/fr/privacybadger.

[5] 2017. uBlock Origin browser extension. (2017). https://www.ublock.org/.

[6] Gunes Acar, Marc Juárez, Nick Nikiforakis, Claudia Díaz, Seda F. Gürses, Frank Piessens, and Bart Preneel. 2013. FPDetective: dusting the web for fingerprinters. In *2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013*, Ahmad-Reza Sadeghi, Virgil D. Gligor, and Moti Yung (Eds.). ACM, 1129–1140. https://doi.org/10.1145/2508859.2516674

[7] M. Ayenson, D. J. Wambach, A. Soltani, N. Good, and C. J. Hoofnagle. 2011. Flash cookies and privacy II: Now with HTML5 and ETag respawning. In *SSRN eLibrary*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1898390.

[8] Frédéric Besson, Nataliia Bielova, and Thomas Jensen. 2013. Hybrid Information Flow Monitoring Against Web Tracking. In *CSF'13*. IEEE, 240–254.

[9] Károly Boda, Ádám Máté Földes, Gábor György Gulyás, and Sándor Imre. 2011. User Tracking on the Web via Cross-Browser Fingerprinting. In *Information Security Technology for Applications - 16th Nordic Conference on Secure IT Systems, NordSec 2011, Tallinn, Estonia, October 26-28, 2011, Revised Selected Papers (Lecture Notes in Computer Science)*, Peeter Laud (Ed.), Vol. 7161. Springer, 31–46. https://doi.org/10.1007/978-3-642-29615-4_4

[10] Yinzhi Cao, Song Li, and Erik Wijmans. 2017. (Cross-)Browser Fingerprinting via OS and Hardware Level Features. In *24th Annual Network and Distributed System Security Symposium, NDSS 2017, San Diego, California, USA, 26 February - 1 March, 2017*. To Appear.

[11] P. Eckersley. 2011. How unique is your browser? *(LNCS)*, Vol. 6205. Springer, 1–18.

[12] P. Eckersley. 2017. The Panopticlick project. (2017). https://panopticlick.eff.org.

[13] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-million-site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi (Eds.). ACM, 1388–1401. https://doi.org/10.1145/2976749.2978313

[14] Gábor György Gulyás, Dolière Francis Somé, Nataliia Bielova, and Claude Castelluccia. 2017. Browser Extension and LoginLeak Experiment. (2017). https://extensions.inrialpes.fr/.

[15] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. 2015. Mitigating Browser Fingerprint Tracking: Multi-level Reconfiguration and Diversification. In *10th IEEE/ACM International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS 2015, Florence, Italy, May 18-19, 2015*, Paola Inverardi and Bradley R. Schmerl (Eds.). IEEE Computer Society, 98–108. https://doi.org/10.1109/SEAMS.2015.18

[16] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. 2016. Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints. In *IEEE Symposium on Security and Privacy, SP 2016, San Jose, CA, USA, May 22-26, 2016*. IEEE Computer Society, 878–894. https://doi.org/10.1109/SP.2016.57

[17] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association.

[18] Jonathan R. Mayer and John C. Mitchell. 2012. Third-Party Web Tracking: Policy and Technology. In *IEEE Symposium on Security and Privacy, SP 2012, 21-23 May 2012, San Francisco, California, USA*. IEEE Computer Society, 413–427. https://doi.org/10.1109/SP.2012.47

[19] Georg Merzdovnik, Markus Huber, Damjan Buhov, Nick Nikiforakis, Sebastian Neuner, Martin Schmiedecker, and Edgar Weippl. 2017. Block Me If You Can: A Large-Scale Study of Tracker-Blocking Tools. In *2nd IEEE European Symposium on Security and Privacy*. Paris, France. To appear.

[20] Nick Nikiforakis, Wouter Joosen, and Benjamin Livshits. 2015. PriVaricator: Deceiving Fingerprinters with Little White Lies. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 820–830.

[21] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2013. Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. In *2013 IEEE Symposium on Security and Privacy, SP 2013, Berkeley, CA, USA, May 19-22, 2013*. IEEE Computer Society, 541–555. https://doi.org/10.1109/SP.2013.43

[22] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2013. Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. In *IEEE Symposium on Security and Privacy*. 541–555.

[23] Lukasz Olejnik and Claude Castelluccia. 2014. Analysis of OpenX-Publishers Cooperation. In *7th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2014)*.

[24] Lukasz Olejnik, Minh-Dung Tran, and Claude Castelluccia. 2014. Selling off User Privacy at Auction. In *21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, California, USA, February 23-26, 2014*.

[25] Mathias Panzenböck. 2012. Social Share Privacy. (2012). http://panzi.github.io/SocialSharePrivacy/.

[26] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and Defending Against Third-Party Tracking on the Web. In *Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2012, San Jose, CA, USA, April 25-27, 2012*, Steven D. Gribble and Dina Katabi (Eds.). USENIX Association, 155–168. https://www.usenix.org/conference/nsdi12/technical-sessions/presentation/roesner

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

[28] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. 2010. Flash Cookies and Privacy. In *AAAI Spring Symposium: Intelligent Information Privacy Management*.

[29] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. 2010. Flash Cookies and Privacy. In *AAAI spring symposium: intelligent information privacy management*. 158–163.

[30] Dolière Francis Somé, Nataliia Bielova, and Tamara Rezk. 2017. Control What You Include! Server-Side Protection against Third Party Web Tracking. In *International Symposium on Engineering Secure Software and Systems (ESSoS)*. To appear.

[31] Ting-Fang Yen, Yinglian Xie, Fang Yu, Roger Peng Yu, and Martin Abadi. 2012. Host Fingerprinting and Tracking on the Web: Privacy and Security Implications. In *Proc. of 19th Annual Network and Distributed System Security Symposium (NDSS)*. Internet Society.

FOLEON EX1018
IPR2026-00428