In Search of Lost Anonymity

Prof. Dr. Gerd Beuster
Waldemar Husser (winf100492)
IT Security Seminar
SS2019 – 22.05.2019

Fachhochschule Wedel

1. Introduction

There are different ways to identify and track a user, with or without their knowledge. In most cases it is enabled through the inherent way the web still works today. To guarantee an ideal viewing experience an HTTP request sends with its header information which is incidentally enough to make the potential visitor uniquely identifiable on the server side of the requested webpage [3]. Tracking through 3rd-party JavaScripts or cross-domain tracking by 3rd-parties is not only bothersome for the privacy concerned user, but also a necessity for the modern browsing experience of knowing what the user wants before they do. Supercookies make browser-independent tracking possible [2] and canvas fingerprinting [1] use the uniqueness of image rendering by the user-machine to identify users. By combining those different data-sets, it is possible to recognize recurring users on a website or follow their web experience across domains.

In this paper I will try to answer the questions posed in chapter 3 of the Red Book [4]. I will look at research and solutions to provide more data control of users and companies, how anonymity in a personalised web can be provided and try to answer the question whether anonymity and personalisation are in conflict.

2. Research

After the implementation of the GDPR (General Data Protection Regulation) [38] on 25 May 2018 websites have to ask the user for permission to use their data first. Now on every domain permission has to be given and in many cases it is not clearly stated how exactly it is being used – are personalised advertisements part of the functionality of a webpage; do you have to give consent to being tracked throughout the web just for using a service? Not giving consent will in some cases deny access entirely, for example to some media outlets.

With GDPR user-consent can be revoked for present and future purposes (Art. 7 GDPR). Section 3 of GDPR “Rectification and erasure” (Art. 16-20 GDPR) enables users to request deletion or correction of any personal data from databases of any service-provider which has or might have those. It defines data should be handled in relation to the purpose they are processed in; only relevant data should be collected (data minimisation), personalised data should be pseudonymised and if not direct identification is not necessary it should be anonymized (Art. 5 GDPR). Users should also be notified if data breaches occurred (Art.33 GDPR). All of the requests for more data-control which the researchers of the Red Book posed are now enforced by law in the EU.

2.1. Anonymity

There are different tools online to see which information your browser makes accessible without you explicit permission and also show how unique your browser is, for example webkay [5], AmIUnique [6], Panopticlick [7].

Incognito/Private browsing only prevents local data access to data placed in the normal session. Cookies in the current private session are only hold during that time or not at all, depending on the browser and settings [8]. Private/Incognito mode of any popular browser does not provide anonymity by itself. With the addition of certain extensions, and correct settings of those, anonymity from third parties could be achieved. For absolute anonymity from trackers the blocking of JavaScript especially from 3rd-party domains is necessary and even blocking of 3rd-party frames. This in return will break the functionality of most modern web-content, since those are highly dependent on JavaScript frameworks or are considered WebApp’s. Simple text representation is still guaranteed unless it is generated by JS.

Most popular choices are NoScript [9], uBlock Origin[10] and uMatrix [11]. Those will block most if not all tracking if setup correctly. A new and less intrusive alternative is the Privacy Badger [12]. For guaranteed https on every visited page the extension “Https Everywhere”[13] is also recommended.

While Firefox is a secure enough browser, it still collects telemetry data [14], and so does Google chrome/Chromium [16]. The Firefox-forks Waterfox [17] and Pale Moon [18] removed telemetry entirely, as did the chromium implementation Iridium [19].

There are browsers which promise security and privacy by design. The Epic Browser integrates most features necessary for anonymity from trackers [20], but it breaks functionality on some sites without an option for compromise.

The best choice for anonymity is using software implementing the Tor protocol [37], for example the Tor Project, which stands in conflict with the personalised web by design [22].

2.1.1. How the Tor Network works

The client(onion-proxy) downloads a signed list of trusted Tor-server from the directory authority. The public keys of those servers are supplied in the source code of the Tor Network, which allows for validation of their authenticity. After the list is received, the client chooses a random connection over the network. The client establishes an encrypted connection to the first Tor-server, the entry node, from there on it keeps establishing encrypted connections to other Tor-servers, until at least three Tor-server are involved. Over this route the data is being transmitted and the last server acts as the exit-node. Unless the network is compromised, only the exit-node can potentially witness the data-stream sent from some client to a service, which can be prevented using end-to-end-encryption.

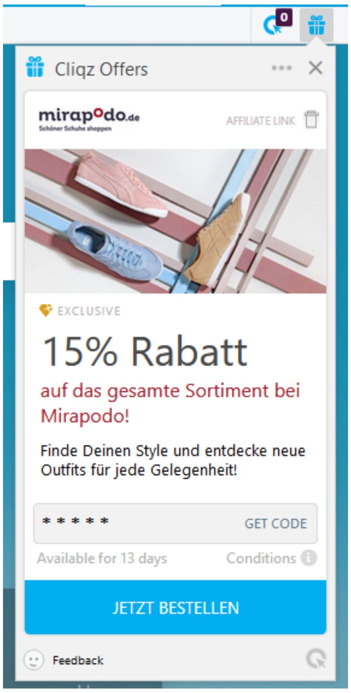
2.2. Personalised Anonymity

A new direction of browser philosophy is to create user-profiles locally on the client side without identifying the user externally. The Cliqz-browser [15] and Brave browser [21] allow similar privacy and security features as the listed extensions above, while still allowing the user to opt-in to some form of personalised content.

2.2.1. Cliqz Browser

Cliqz MyOffrz GmbH uses a browser with integrated data collection, analysis and marketing. Their MyOffrz Software, integrated in the Cliqz-browser, downloads all available offers from their partners.
No user-identifying and user-behaviour data is sent to any party. The decision to display an offer is made locally in the browser by its learning algorithm [23, 24].

All available offers by MyOffrz are loaded to the browser and only the algorithm in the browser decides when and if to show an offer. The user behaviour is analysed with conditional trees, by looking at search terms and if a condition is met, a fitting offer will be available. They promise that no personalised data is sent to any party - not even themselves, which they achieve by obfuscating the user IP in static data reports via proxy networks.
Corporations, which want to partake in the offers-system, have to become partners with MyOffrz. The offers are displayed as a notification in the browser toolbar (Fig. 1), and if a user clicks on an offer the browser-company gets a provision. [34, 35, 36]



Cliqz Offers

mirapodo.de
Schöner Schuhe shoppen

AFFILIATE LINK

EXCLUSIVE

15% Rabatt

auf das gesamte Sortiment bei Mirapodo!

Finde Deinen Style und entdecke neue Outfits für jede Gelegenheit!

Fig. 1. Screenshot of the offer notification in the browser-toolbar.

2.2.1. Brave Browser

Brave’s algorithm rewards user attention on websites and user-advert interaction with their ethereum based token, BAT [25], but only if users choose to partake [26]. The condition for publishers (webdomains) [28] and advertisers [27] to partake in this form of exchange is to actively partner with Brave.

Cliqz and Brave have two different approaches to personalised anonymity and I believe in the future marketing, user analytics and user suggestions will be done mainly by the used browser themselves. It is not far fetched to believe, that such a feature might be added to Mozilla Firefox in the future, since they have partnered once before with Cliqz GmbH [29, 30].

2.3. Handling of personal data

Answering the example problems has become rather difficult. The problems posed by the researchers of the red book concerning personal data and privacy in the web as they existed in 2013 have changed since and thus – I believe – made some of the suggested approaches obsolete. With implementation of GDPR data concerns are not a nice to have, but enforced by law EU-wide, and since many globally-acting non-EU companies do not want to lose out on that market, they also abide by those laws.

2.3.1. Example Problem: Provide personal data in a provably k-anonymous form

The suggestion (see quote below) posed here was to change how applications ask for authentication of user attributes – like age – by requesting data sets, which also include the user, to hide the identity of the requesting user from the application doing the request.

> "*In this research one might change the model and force applications to ask data from sets of users. The set will provide data […], without revealing, however, which individual user of the set is using the current application.*"
> - Quote from page 25 of the Red Book [4].

Neither have I found research for this, nor can I imagine how this might be in anyway useful.

2.3.2. Example Problem: Privacy in an eponymous world

The goal of this problem (Chapter 3.6, page 25, the Red Book [4]) was to find a solution to obfuscate a users true interests, while they are logged in with – I presume – a social networking service account, like google, facebook, myspace or twitter, and are browsing the web on specific pages, which would represent their interests.
It was asked if there was a way to automate the process of confusing classification algorithms of social networks, which are tracking user behaviour by design. I.e if the user visits the website of Fox News, an algorithm would then visit other websites, like MSNBC, to obfuscate the users true interests from the social network providers.

While this is an interesting approach of hiding one's interests, the realisation of such a process would imply a tool with a trained classification algorithm itself, so it can automatically search the other views. The

algorithm would have to be smart enough to fool a far smarter opponent like the google search engine.
I have not found any tool or research to achieve such a thing, which is understandable since going anonymous in the Tor Network or using a VPN and private browser is far easier to achieve the same goal of keeping one's true interests hidden.

2.3.3. Example Problem: Honey-profiles

The suggested research (Chapter 3.6, page 25, the Red Book [4]) was to fill the databases of different websites and applications with fake profiles with distinctive and unique features, so that leakage of profiles could be recognized and followed. Those profiles would be whistles to the party involved with care of the personal data. Since every website/application would have their own fake profile, each of them would be able to recognize their own leaks. There is no need for research in this regard; fake data entries to recognize breaches are not a new development:
Honeytokens [33] are fake data entries, which should not be accessed by anyone. Every access implies a data breach in some form.

3. Conclusion

The fears concerning lost anonymity in the web are a valid concern. With the implementation of the GDPR the respect for personal data became not only relevant for the EU-public, but also globally. Other nations are following step – in their own time – and global companies comply willingly or by force [31, 32]. Users are now more interested in alternatives to the known web of constant tracking and I believe that the successful emergence of browsers, which are designed around a local user-behaviour-analysis, is a sign that the movement of the deciding agent for advertising will move away from external third parties to local algorithms in the coming future.

While testing the suggested browsers I have come to the conclusion, that personal data does not have to be forfeited to gain a personalised web experience. For now the added value is relatively low, since not many companies partnered with Cliqz or Brave. But I believe similar browsers will emerge with different design philosophies on how that synergy between personalisation and anonymisation could be achieved.

4. Used references

[1] Keaton Mowery and Hovav Shacham, Pixel Perfect: Fingerprinting Canvas in HTML5.
URL: <https://hovav.net/ucsd/dist/canvas.pdf>, date of access 10.05.2019

[2] Seth Schoen, New Cookie Technologies: Harder to See and Remove, Widely Used to Track You.
URL:
<https://www.eff.org/deeplinks/2009/09/new-cookie-technologies-harder-see-and-remove-wide>, date of access 10.05.2019

[3] Peter Eckersley, Browser Versions Carry 10.5 Bits of Identifying Information on Average.
URL: <https://www.eff.org/deeplinks/2010/01/tracking-by-user-agent>, date of access 10.05.2019

[4] sysec, The Red Book - A Roadmap for Systems Security Research, Chapter 3.
URL: <http://www.red-book.eu/m/documents/syssec_red_book.pdf>, date of access 10.05.2019

[5] Robin Linus, webkay.
URL: <https://webkay.robinlinus.com>, date of access 10.05.2019

[6] AmIUnique.
URL: <https://amiunique.org>, date of access 10.05.2019

[7] EFF, Electronic Frontier Foundation, Panopticlick.
URL: <https://panopticlick.eff.org>, date of access 10.05.2019

[8] Chris Hoffman, How Private Browsing Works, and Why It Doesn't Offer Complete Privacy.
URL:
<https://www.howtogeek.com/117776/htg-explains-how-private-browsing-works-and-why-it-doesnt-offer-complete-privacy/>, date of access 10.05.2019

[9] Giorgio Maone, NoScript.
URL: <https://addons.mozilla.org/en-US/firefox/addon/noscript/>, date of access 10.05.2019

[10] uBlock Origin.
URL: <https://github.com/gorhill/uBlock>, date of access 10.05.2019

[11] Raymond Hill, uMatrix.
URL: <https://github.com/gorhill/uMatrix>, date of access 10.05.2019

[12] EFF, Electronic Frontier Foundation, Privacy badger.
URL: <https://www.eff.org/de/node/99095>, date of access 10.05.2019

[13] EFF, Electronic Frontier Foundation, Https Everywhere
URL: <https://www.eff.org/de/https-everywhere>, date of access 10.05.2019

[14] Effectively Measuring Search in Firefox
URL: <https://blog.mozilla.org/data/2018/08/20/effectively-measuring-search-in-firefox/>, date of access 10.05.2019

[15] Cliqz GmbH, Privacy Policy - Cliqz.
URL: <https://cliqz.com/privacy-browser>, date of access 10.05.2019

[16] Google Chrome Privacy Notice
URL: <https://www.google.com/intl/en/chrome/privacy/>, date of access 10.05.2019

[17] Alex Kontos, Adam Wood, Waterfox.
URL: <https://waterfoxproject.org/en-US/>, date of access 10.05.2019

[18] Moonchild Productions, Pale Moon.
URL: <https://www.palemoon.org/>, date of access 10.05.2019

[19] Open Source Business Alliance e.V., Iridium.
URL: <https://iridiumbrowser.de/>, date of access 10.05.2019

[20] Hidden Reflex Inc., Epic Browser.
URL: <https://www.epicbrowser.com/>, date of access 10.05.2019

[21] Brave Software Inc., Brave.
URL: <https://brave.com/>, date of access 10.05.2019

[22] Mike Perry, Erinn Clark, Steven Murdoch, Georg Koppen. The Design and Implementation of the Tor Browser.
URL: <https://2019.www.torproject.org/projects/torbrowser/design/>, date of access 10.05.2019

[23] Cliqz MyOffrz GmbH, myoffrz for Users.
URL: <https://myoffrz.com/fuer-nutzer/>, date of access 10.05.2019

[24] Cliqz GmbH, Cliqz introduces MyOffrz, a completely new form of advertising.
URL:
<https://cliqz.com/en/magazine/press-release-cliqz-introduces-myoffrz-a-completely-new-form-of-advertising>, date of access 10.05.2019

[25] Brave Software Inc., BAT Whitepaper.
URL: <https://basicattentiontoken.org/BasicAttentionTokenWhitePaper-4.pdf>, date of access 10.05.2019

[26] Brave Software Inc., BAT User Terms of Service.
URL: <https://basicattentiontoken.org/user-terms-of-service>, date of access 10.05.2019

[27] Brave Software Inc., BAT Advertiser Terms of Service.
URL: <https://basicattentiontoken.org/advertiser-terms-of-service/>, date of access 10.05.2019

[28] Brave Software Inc., BAT Publisher Terms of Service.
URL: <https://basicattentiontoken.org/publisher-terms-of-service/>, date of access 10.05.2019

[29] Ein neues Cliqz-Experiment in Firefox.
URL: <https://blog.mozilla.org/press-de/2017/10/06/ein-neues-cliqz-experiment-in-firefox/>, date of access 10.05.2019

[30] Charlie Osborne, Mozilla pilots Cliqz engine in Firefox to slurp user browsing data.
URL:
<https://www.zdnet.com/article/firefox-tests-cliqz-engine-which-slurps-user-browsing-data/

[31] Charlie Osborne, Facebook could face $1.63bn fine under GDPR over latest data breach.
URL:
<https://www.zdnet.com/article/facebook-could-face-billions-in-fines-under-gdpr-over-latest-data-breach/>, date of access 10.05.2019

[32] Adam Satariano, Google Is Fined $57 Million Under Europe’s Data Privacy Law.
URL: <https://www.nytimes.com/2019/01/21/technology/google-europe-gdpr-fine.html>, date of access 10.05.2019

[33] Lance Spitzner, Honeytokens: The Other Honeypot
URL: <https://www.symantec.com/connect/articles/honeytokens-other-honeypot>, date of access 10.05.2019

[34] Cliqz GmbH, “MyOffrz - Teil 1: Smarter surfen mit maximaler Privatsphäre”.
URL:
<https://cliqz.com/magazine/myoffrz-teil-1-smarter-surfen-mit-maximaler-privatsphaere>, date of access 20.09.2019

[35] Cliqz GmbH, “MyOffrz – Teil 2: Zielgerichtete Angebote und Privatsphäre sind kein Widerspruch”.
URL:
<https://cliqz.com/magazine/myoffrz-teil-2-zielgerichtete-angebote-und-privatsphaere-sind-kein-widerspruch>, date of access 20.09.2019

[36] Cliqz GmbH, “MyOffrz – Teil 3: Wie funktioniert die MyOffrz-Technologie im Detail?”.
URL:
<https://cliqz.com/magazine/myoffrz-teil-3-wie-funktioniert-die-myoffrz-technologie-im-detail>, date of access 20.09.2019

[37] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The Second-Generation Onion Router
URL: <https://svn.torproject.org/svn/projects/design-paper/tor-design.html>, date of access 20.09.2019

[38] Official Journal of the European Union, REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>, date of access 10.05.2019