UNIVERSITY OF TARTU
Institute of Computer Science
Computer Science Curriculum

Magnus Valgre

Tracking And Privacy: The Case of News Site Delfi

Bachelor's Thesis (9 ECTS)

Supervisor: Arnis Paršovs, PhD

Tartu 2021

Tracking And Privacy: The Case of News Site Delfi

Abstract: Privacy and tracking on the internet are concerns that have gotten more and more attention over the last few years. Among the biggest perpetrators of online tracking are news sites. Since many provide their content for free, and do not have an external funding source, they need to monetize pageviews by displaying advertising. The purpose of this thesis is to provide a privacy analysis of the Estonian news site delfi.ee. Delfi was chosen because it is the largest and most visited news site in Estonia. During the research some privacy issues were found showing that Delfi is currently not GDPR compliant. This thesis provides an overview of some commonly used tracking techniques, how they apply in the context of Delfi, and an analysis of Delfi's privacy policies.

Keywords:

Tracking, privacy, cookies, HTTP requests, social login, local storage, privacy policy, cookie policy, GDPR

CERCS: P170 Computer science, numerical analysis, systems, control

Jälgimine ja privaatus: Delfi uudisteportaali uuring

Lühikokkuvõte: Privaatsus ja jälgimine internetis on mured, mis on saanud viimastel aastatel aina rohkem tähelepanu. Ühed kõige suuremad jälgimise läbiviijad on uudisteportaalid. Kuna paljud pakuvad oma sisu tasuta ja neil puuduvad välised raha-allikad, peavad nad raha teenima külastajatele reklaami näitamisega. Käesoleva uurimustöö eesmärgiks on läbi viia privaatsusanalüüs Delfi uudistelehele. Delfi valiti uurimustöö aluseks selle tõttu, et tegu on Eesti mahukaima ja suurima külastajate arvuga uudisteportaaliga. Uurimustöö tulemusena leiti mõned privaatsusriskid, mille tõttu Delfi hetkel ei ole kooskõlas Euroopa Liidu isikuandmete kaitse üldmäärusega. See uurimustöö annab ülevaate mõningatest levinud jälgimistehnikatest ja kirjeldab, kuidas neid kasutatakse Delfi uudisteportaali kontekstis. Lisaks analüüsitakse ka Delfi privaatsuspoliitikat.

Võtmesõnad:

Jälgimine, privaatsus, küpsised, HTTP päringud, sotsiaalmeedia kaudu sisse logimine, kohalik ladustamine, privaatsuspoliitika, küpsiste poliitika, ELIKÜM

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine

Table of Contents

List of Abbreviations

API – Application Programming Interface

FB – Facebook

FLoC – Federated Learning of Cohorts

GDPR – General Data Protection Regulation

HTML – Hypertext Markup Language

HTTP – Hypertext Transfer Protocol

HTTPS – Hypertext Transfer Protocol Secure

WPM – Web Privacy Measurement

1 Introduction

Privacy and tracking on the internet are concerns that have gotten more and more attention over the last few years. In 2018 the Cambridge Analytica scandal broke and revealed how a company used Facebook to gather data on tens of millions of users that was then used to show targeted political advertising [1]. The same year the European Union enacted its General Data Protection Regulation, the strictest privacy law in the world. The GDPR imposes itself onto organizations all over the world if they want to process information of individuals situated in the European Union [2]. So it is safe to say that privacy is on the public mind.

Among the biggest perpetrators of online tracking are news sites. Since many provide their content for free, and do not have an external funding source, they need to monetize pageviews by displaying advertising [3]. More advertising means more potential trackers.

The purpose of this thesis is to provide a privacy analysis of the news site Delfi[1]. Delfi was chosen because it is the largest and most visited news site in Estonia. Among websites covered by the Gemius traffic analysis tool, it is also the most popular site overall with 712 000 visitors in March of 2021 [4].

The thesis provides an overview of some common privacy issues encountered on the internet and then analyzes them in the context of Delfi. The thesis also provides an analysis of the privacy and cookie policies of Delfi's parent company Ekspress Meedia AS.

The thesis is structured as follows. Section 2 contains a brief overview of privacy in the World Wide Web, why it is important, and how Estonians feel about their data being gathered. Section 3 contains the overview of the tracking techniques and the analysis of how these techniques are present in Delfi's website. The final section is the conclusion. The appendix contains a list of cookies that Delfi, and third-party partners of Delfi, set on the user's computer during testing. Where reliable information was available, a description of the cookie's purpose is also given.

[1] https://www.delfi.ee

1.1 Privacy and tracking on news sites

Web tracking is the process of gathering information on users for the purpose of identifying them across different websites and building a list of visited web pages. This is usually done by third parties like advertisers, social media widgets or analytics engines. Since the 1990s tracking on the World Wide Web and the methods used have evolved greatly. Tracking on the internet has become more prevalent and complex, and there are now more trackers that exhibit different types of behaviours. Over time, web tracking has also become more concentrated, meaning that the same trackers cover an increasing number of websites. In 2016 Google Analytics covered a third of the top websites [5]. By 2022 Google has promised to stop the use of cookies altogether and is looking to replace them with a technology they have named FLoC (Federated Learning of Cohorts) which, according to critics, will concentrate even more of the tracking ecosystem into Google's hands [6].

Privacy is a human right and important to our autonomy and dignity. It helps us protect ourselves from unwanted interference in our lives and allows us to place boundaries on who has access to us, our communications and information [7]. This is also true for our online activities. Privacy on the internet means that we have control over the data that our actions generate. That we know who is gathering this data, for what purpose, and that we can have control over it [8].

In 2020, a study was ordered by the Estonian Ministry of Justice and carried out by Kantar Emor [9]. It came to the conclusion that a majority of internet users find it disturbing when websites gather information on their previous actions to provide personalized advertisements. Furthermore, respondents felt negatively towards general behavior analysis on the internet carried out by social media companies. Generally, younger people are more against it than older people. Respondents in a study conducted in 2014 by researchers from the Estonian Institute of Human Rights answered similarly to questions concerning what actions they find disturbing [10]. So concerns and attitudes towards people's online privacy have remained largely the same from 2014 to 2020.

1.2 Methods and tools

The main way to analyze how a website treats its users' privacy is to use the website going through some normal workflows, while at the same time investigating the network traffic that is exchanged between the user's computer and the server, and the browser's behaviour.

For this purpose we used a combination of the network analyzer WireShark, Firefox Developer Tools, that are a set of web developer tools built into the Firefox browser, and a privacy research framework named OpenWPM.

Wireshark[2] is a free and open-source packet analyzer that is widely used for analysing network traffic. For this study we used Wireshark to monitor network traffic while using the website and to analyze network requests made when the website loads.

OpenWPM[3] is an open-source software that is meant for conducting privacy measurements on the web. It was first developed by researchers at the Princeton University but since 2019 is maintained by Mozilla. It can be used for large scale measurements to crawl millions on websites but can be also beneficial to use on small-scale investigations as this.

Firefox Developer Tools[4] were used in this work to monitor HTTP requests and responses, cookies and local storage.

For this investigation we used OpenWPM to gather data on: exchanged HTTP requests and responses, the cookies set, and JavaScript scripts that were called as the site loaded.

[2] https://www.wireshark.org/

[3] https://webtap.princeton.edu/software/

[4] https://developer.mozilla.org/en-US/docs/Tools

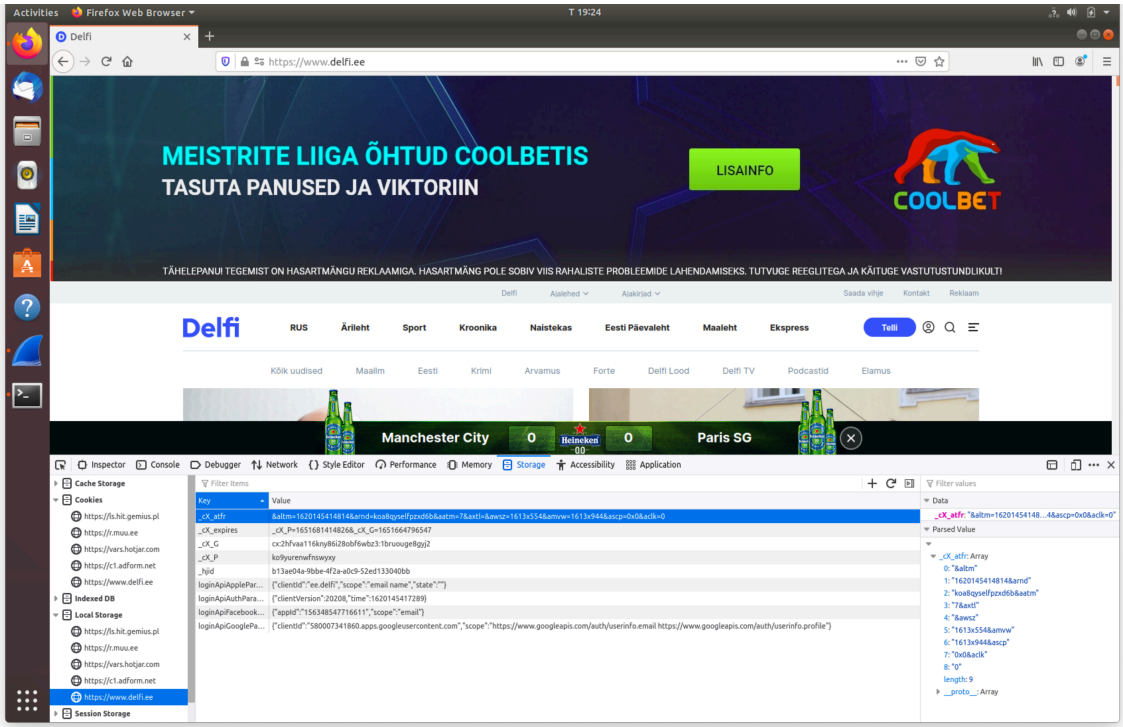

Figure 1. An example of using Firefox's built-in developer tools for privacy analysis. Firefox is displaying the local storage variables set by Delfi's website.

Most of the investigation was done manually using the Firefox browser and its web developer tools. Firefox allows to monitor network traffic and see what data is stored on the client side. It can also be used to test how different privacy enhancing measures work and what effect they have on the user experience. Figure 1 one shows how Firefox's developer tools can be used to find out what data is present in the browser's local storage.

We used Firefox's Development Tools and Wireshark to investigate network traffic while the site loads (especially HTTP requests) and while navigating between different pages. This allowed us to investigate cookie and local storage usage. The Firefox browser was set up with default settings and no privacy enhancing add-ons were installed. We also ran a OpenWPM crawl on Delfi. OpenWPM visits a specific website and records its results in a SQL database that is easy to navigate and can be queried for specific data. Among the data the crawl recorded, the most interesting were HTTP requests and their responses, Javascript scripts that were called and cookies that were set. The crawl allowed us to combine a list of cookies and the domains that set them and also to look for signs of browser fingerprinting.

We tried to look for signs of common canvas fingerprinting techniques but were unable to find any.

To confirm the owners and purposes of the cookies, we used databases provided by other organizations (Web Cookies Scanner[5] and CookieServe[6]). For some of the cookies, documentation was available on the websites of the companies that set and use these cookies.

A privacy and cookie policy analysis was done by following guidelines published by privacy-minded organizations[7] and following methods used in previous investigations [11].

The screenshots included in this work have been made by the author. The diagrams have been created using Diagram Editor[8].

[5] https://webcookies.org

[6] https://www.cookieserve.com

[7] https://gdpr.eu/

[8] https://www.diagrameditor.com/

2 Analysis of user tracking on Delfi's website

In this section, an analysis of the privacy and cookie policies of Delfi's parent company Ekspress Meedia AS is provided, and an overview of some common privacy issues and how they relate to Delfi.

2.1 Privacy policy and GDPR compliance

A privacy policy is a public document that explains how the personal information of users is gathered by an organization and how it is managed. Personal information is any piece of information that can be used to identify a person. For example, a person's name, date of birth, home address, or geographic location [12]. Even an individual's web browsing history can be considered personal information because it can be very unique from person to person [13].

Any company operating in or processing personal information of users in the EU, regardless of their actual location, must comply with the GDPR. Among other requirements, that means having a GDPR compliant privacy policy.

According to the GDPR [12], a privacy policy must be:

- In a concise, transparent, intelligible, and easily accessible form
- Written in clear and plain language
- Delivered in a timely manner
- Provided free of charge

If an organization or a company wishes to use cookies on their website, it must notify the user of their intentions and receive explicit consent. Before consent is received from the user, no cookies that are not strictly necessary for the operation of the website should be set. Prior consent must be received before processing any personal information, but most widely, in the context of the web, this means asking permission to set cookies. Most websites choose to do so by displaying a notifying banner. As an example, Figure 2 shows how this is currently implemented on the homepage of the Institute of Computer Science of the University of Tartu. The user can consent by clicking on the "OK" button.

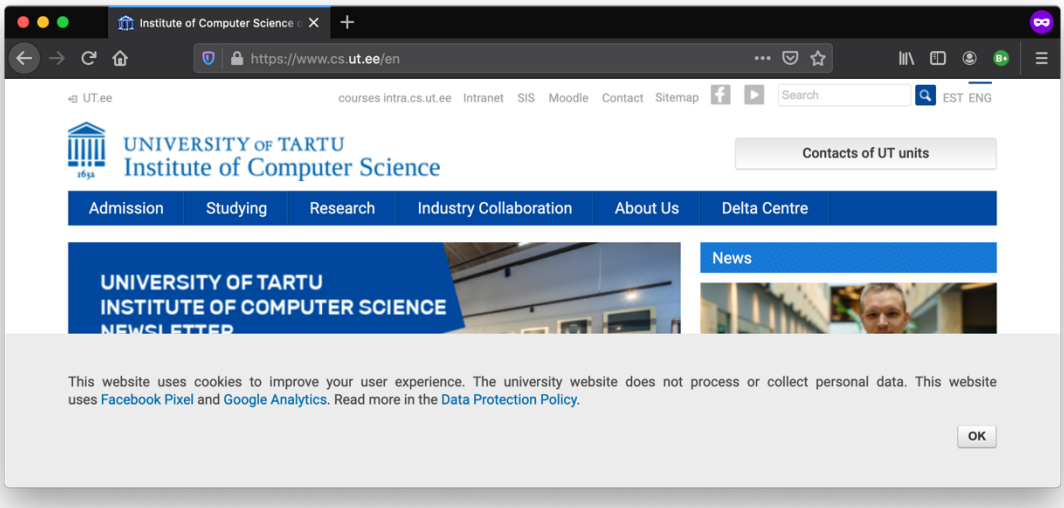

Figure 2. Screenshot taken of the homepage of the Institute of Computer Science of University of Tartu. Bottom of the page exemplifies how a cookie consent notification can be implemented.

Many websites also decide to compile a cookie policy separate of the privacy policy that users can review before giving consent.

According to the GDPR [14], a cookie policy should:

- Describe the type of cookies used
- Describe the purpose of the cookies
- Indicate all third parties that install or could install cookies
- Be available in all languages the service is provided in

In addition to correctly informing the user, the website operator must also document and store the received consent, and make it easy to withdraw if the user should change their mind. The service should be accessible even if the user refuses to give consent [15].

Delfi is a part of the Ekspress Meedia media company that has several print publications under its umbrella besides the Delfi news site. All of these print publications also have a online version as a subdomain to Delfi [16]. Delfi shares its privacy policy [17] with the other Ekspress Meedia publications. In addition to the privacy policy, there is also a separate cookie policy [18].

Ekspress Meedia's privacy policy defines all the involved parties (the subject, the controller and the processor), what is personal data, for what purpose it is gathered, and what principles are followed in processing. Personal data is any data that helps to identify a person. Such as a name or contact information.

A list of partners[9] is also provided who are allowed by Ekspress Media to process users' personal data (for example, Google, Facebook and Adform).

In the list of types of data gathered it is mentioned that "web identifiers, such as the IP-address of your device when you visit our website" are collected. This sentence leaves the question what other web identifiers might be collected. Other words and phrases like "mostly" and "among other things" are also used that make the list feel incomplete.

The privacy policy also states that Ekspress Meedia, with the help of aforementioned partners, will profile the subject and summarise behaviour patterns in order to develop targeted advertising profiles. The policy claims no individual profiles are created and all behaviour patterns are summarised and categorized into target groups. Furthermore, that membership in any target group is not recorded by Ekspress Meedia and no individual's target groups can be determined.

The policy seems to be in line with the articles 12, 13 and 14 of the GDPR which cover how a privacy notice should be created [12].

Ekspress Meedia's cookie policy states in clear terms that they use cookies and, since they serve content from third parties on their website, some of them might be third-party cookies.

The policy continues with informing the user that cookies can be disabled by changing browser settings and that different browsers have different methods to achieve this, although disabling cookies can cause issues with the website. Inexplicably, for more information, the policy links to the cookie policy of the Estonian Data Protection Inspectorate[10].

The policy then gives an introduction to what cookies are and what types of cookies are used on Ekspress Meedia sites, but again links to the Data Protection Inspectorate's homepage for more information. The information there does not add anything new to the information provided in Ekspress Meedia's cookie policy.

[9] https://www.ekspressmeedia.ee/partnerile/andmekaitse/volitatud-tootlejad/

[10] https://www.aki.ee/et/kupsised

Ekspress Meedia's cookie policy is pretty standard and is available in English, Estonian and Russian. The only thing that is missing is a list of third-party partners that might be setting cookies on behalf of Ekspress Meedia, but this is list available under the overall privacy policy.

At the time of this writing, Delfi was missing a cookie notice for first time users. Under the GDPR, cookies are form of identifiers [19], and as such, need the website operator to receive explicit consent from the user before setting any that are not strictly necessary. Since Delfi's website does not ask for consent before setting third-party cookies, as the cookies are loaded along with the rest of the site, currently Delfi is not GDPR compliant.

Using the Internet Archive's Wayback Machine[11], we were able to determine that as of March 17th of 2021[12], the cookie consent banner does not appear. We reached out to Ekspress Meedia for comment on this and they replied that they are aware of the issue. Delfi moved to a new platform in the middle of March and there are still some bugs that need to be fixed. They stated that hopefully this issue will be fixed in the first week of May.

The Wayback Machine allowed us to test if Delfi sets any cookies before the user gives their consent and we were able to determine that the website indeed sets some third-party cookies. This means that Delfi was not GDPR compliant even when the banner was shown to users. However, the Wayback Machine does not recreate the webpage and its back-end perfectly, so this cannot be said with absolute confidence. It is possible that there are some resources that block the loading of cookies on the regular site that are not present in the archived version.

Under the GDPR, there have been fines imposed for companies operating in the European Union for failing to properly receive the user's consent [20], with France fining Amazon [21] and Google [22] 35 and 100 million euros respectively. Estonian data protection authorities have not yet employed such fines. By the beginning of 2021, Estonian authorities have imposed fines in the sum of 408 euros, choosing instead to warn offenders with penalty warnings [23]. One of the possible reasons for this is the relative legal complexity of imposing fines in Estonia compared to other EU countries. According to Estonian law, fines cannot be imposed under administrative procedure and have to be processed as a

[11] https://archive.org/web/

[12] https://web.archive.org/web/20210316060223/https://www.delfi.ee/

misdemeanour. Misdemeanours require the accused party to be proven guilty beyond any doubt and, before any fines can be levied against an organization, a natural person has to be found guilty of the misdemeanour. This makes it difficult to impose fines on organizations for breaching the GDPR. The second reason is the passivity of the Data Protection Inspectorate [24].

Along with the privacy policy, Ekspress Meedia provides a list of third-party partners who are authorized to process users' information. However, some of the companies associated with the set cookies do not appear on the list. The companies that have not been listed in the privacy policy, but set cookies, are: Clickonometrics, Bidtheatre, HotJar, BidSwitch and AppNexus. Although this is not a direct violation of the GDPR. The GDPR states that a website must provide a list of all processors and sub-processors if asked by the data subject but such a list does not have to be readily available [25]. The processors must get authorization from the owener of the website before contracting any sub-processors [26].

Another possible privacy issue might be the type of connection a website uses to communicate. An unsecured HTTP connection could pose a security and privacy risk to users. An unsecure connection would mean that network packets between the browser and the server are unencrypted and thus vulnerable to eavesdropping and man-in-the-middle attacks [27]. Delfi avoids this issue and uses a secure HTTPS connection by default and the certificates are up to date. If a user tries to access the site via HTTP, they will be automatically redirected to the HTTPS site.

2.2 Cookies

The most well-known tracking technique is the use of third-party cookies. Cookies are small text strings that contain key-value pairs. The cookie file is created when a browser requests a resource from a server. The server replies to the request and sends the resource. Along with the resource, the server can also send a cookie in the HTTP response header. If a request to the same server is made again, then the cookie, that was previously set, is sent back in the header of the new HTTP request. When the server receives the request, it will recognize the user by the identifier contained in the cookie. This means that the server is able to follow around a user as they navigate around the website [28].

Cookies are connected to the domain that set them. If the domain is the same as the one the browser is currently on, then the set cookie is a first-party cookie. First-party cookies are

used to enhance the website's functionality, usually to personalize or manage a session (keeping track of a shopping cart or login information).

A website can also display content that is hosted under other domains than the one the browser is currently on. Third-party cookies are set by such content that is hosted under other domains. For example, a user is browsing the site `example1.org`, which displays advertisements hosted on `ad.tracker.com`, and one of the requests to `ad.tracker.com` to fetch an advertisement sets a cookie for the domain `ad.tracker.com`. Later on, the user is browsing the site `example2.org` and this site serves advertisements hosted by `ad.tracker.com` as well. When the browsers sends a request to `ad.tracker.com` while on the site `example2.org`, the cookie that was set on the first site is sent to the server in the HTTP request. This way `ad.tracker.com` can recognize a user across multiple sites [29].

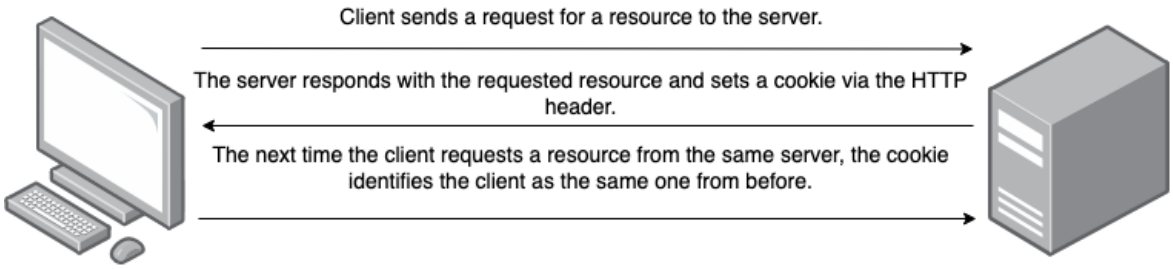

Fig 3. How a HTTP cookie works.

As of March 2021, Google, the creator of the Chrome browser, has announced it is stopping support of third-party cookies by the year 2022 [30]. Chrome, which is the most popular browser in the world on mobile and desktop computers [31], is the last of the big three browsers behind Mozilla's Firefox and Apple's Safari to do so [32], which disable third-party cookies by default.

During the testing Delfi set a total of 45 different cookies onto the user's computer. Of the 45 cookies 24 were first-party and 21 third-party.

A table listing the different cookies set on the user's computer when visiting Delfi can be found in the appendix.

Since cookies themselves do not provide any information on how they are used and are only key value pairs of strings, the only hints to their purpose are usually their names and the

domains that set them. We have used cookie databases provided by other organizations to help determine the purpose of these cookies.

At least 13 of the set cookies were advertisement cookies. Meaning they are used by different providers to show users advertisements and marketing campaigns. These are the cookies that can track a user across different websites and collect information on user behaviour.

Advertisement cookies were linked to the following companies: Google, Facebook, Adform, Bidswitch, Appnexus, Bidtheatre, DoubleClick (acquired by Google in 2007).

At least 5 of the set cookies were Analytics cookies. These cookies are used to gather information on how users interact with the website, amount of visitors, where the traffic is coming from.

Analytics cookies were linked to the following companies: Google, Cxense and HotJar.

The number of cookies set by Delfi is noticeably higher than other top Estonian news sites. According to captured network traffic, Postimees sets 13 cookies and ERR sets 17 cookies, while Delfi sets a total of 45 cookies. According to Web Cookie Scanner's[13] statistics, the average number of cookies that a site sets is 5. Due to heavy reliance on advertisement content, news sites use considerably more third-party cookies than other sites. In a study done across 7 EU countries, it was found that the average number of third-party cookies a news site sets is 81 [33].

If all third-party cookies are disabled in the browser, Delfi still remains usable and does not force the user to enable them in order to keep accessing the content. This means that cookie settings and differences between major browsers do not affect user experience, since Chrome is the only major browser that still allows third-party cookies by default. If all cookies are disabled, then some functionality is lost, like logging in through Facebook's social login API.

2.3 Tracking pixels

Tracking pixels are also known as a web beacons, web bugs, tags, or pixel tags. It is a tiny 1x1 pixel image that is loaded along with the rest of the page and remains invisible to the user. When the request for the image is sent to the server, the server logs and timestamps

[13] https://webcookies.org/number-of-cookies/

the request. Tracking pixels are usually used to count visitors and ad impressions (an impression is counted each time and ad is shown) on a site by analytics providers [34].

Along with the request for the image file, the browser passes additional information to the server via HTTP headers. The information typically includes the IP address of the user's computer, timestamp of the request and the provider and the user's browser version.

Another common usage is to embed a tracking pixel into an email containing HTML content. When the email is loaded, the tracking pixel is also loaded and the server hosting the content is made aware of this. This way the sender of the email can verify if the recipient actually opened the email. The sender of the email also receives a timestamp and the IP address of the user's computer. This means the sender has a rough estimate of the recipient's location and the time the email was opened [35].

Delfi uses tracking pixel solutions provided by the companies Adform and Facebook.

```
<img src="https://track.adform.net/
adfserve/?bn=***;1x1inv=1;srctype=3;ord=***"  border="0"  width="1"
height="1"/>
```

Figure 4. An HTML image tag containing Adform tracking pixel code. Example captured with an OpenWPM crawl.

Figure 4 shows an example of an HTML image tag containing the code that loads an Adform tracking pixel. The `bn` parameter contains Adform's tracking point ID. This is a unique identifying code that helps AdForm's servers identify what advertisement was loaded and count it as an impression.

The `ord` parameter is called a cache buster. To save time and bandwidth, browsers cache images and other resources onto the hard-drive so that they would not have to be downloaded again if they are needed for a second time. This also works with advertisements and tracking pixels. If the user browses different pages on the same website that contain the same advertisement, the browser would just display the same one it has cached again. Since the server would not receive a request for the pixel, the impression would not be counted. This is counteracted by adding a random parameter to the tracking pixel's URL each time it is loaded to trick the browser into thinking it is a new resource [36].

Facebook's tracking pixel is different in that it is actually a piece of JavaScript code that is called each time a user takes an action. The user's actions are noted and the data is forwarded to Facebook. The script can differentiate between a wide array of events. For example, when a purchase is made, a user registers on a site, or something is added to a shopping cart [37]. In Delfi's case, the action seems to be a simple page view. Meaning, when the page is loaded, a URL under Facebook's domain is called and Facebook's analytics logs that this page has been viewed by a user.

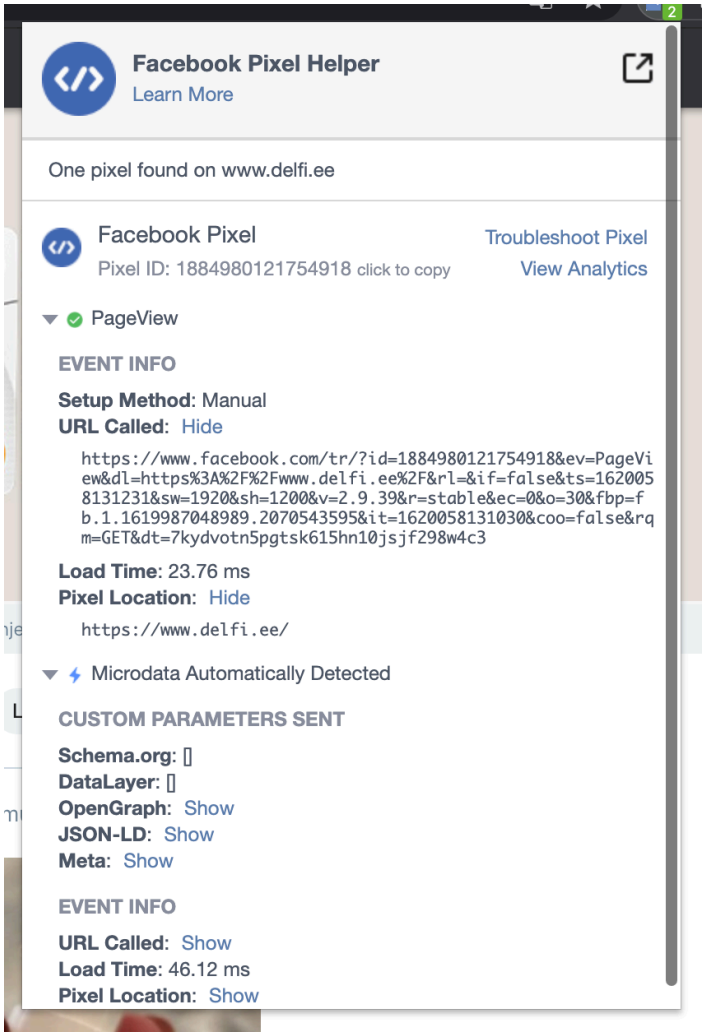

Figure 5. Image of the Facebook Pixel Helper showing info on the tracking pixel on Delfi's website.

Figure 5 shows a screenshot of the Facebook Pixel Helper, a Chrome plugin by Facebook that can help web developers troubleshoot their implementations of the Facebook Pixel, but can also be used by researchers to easily see if and how a website uses them.

This tracking pixel is the main tool Facebook uses to track users across the web to gather data on their browsing habits. The operators of a website set up a Facebook Pixel in order to gain access to Facebook's analytics tools. They can use the tools to gather insights into how users interact with their site and in turn, Facebook can also use the data to improve the accuracy of targeted advertisements [38].

The way tracking pixels set cookies is very similar to the way regular third-party cookies are set (see Section 2.2).

2.4 Social login

Users can register an Ekspress Meedia account to purchase access to articles behind a pay-wall or an advertisement-free version of the website. To make user registration easier, it is possible for the user to create an Ekspress Meedia account by authenticating their identity using their Facebook, Google or Apple account. This is known as social login and is increasingly popular due to the convenience it offers. Instead of creating a new account, the user can connect an existing account that the user probably already has, and start using a new service with less friction. This convenience can come at a cost to privacy. If a user does not check carefully what data they are allowing the social login authorizer to share with the website, they can end up sharing more than they were planning on [39].

Delfi only requests the user's full name, profile picture and e-mail address from the authorizer (i.e., Facebook, Google or Apple) when signing up through social login, which is the reasonable minimum amount of information.

Even if a website is acting responsibly and does not request excessive information from the authorizer, the other side is still that the user is giving more information about themselves to the social login provider. In this case the user would be providing Facebook, Google or Apple with the information that they are now registered users of the Delfi website. Since Delfi is also using Facebook's Pixel, logging in through Facebook will enable it to directly connect your identity and browser to the web pages you view on Delfi. Even if a user does not log in through Facebook but has a Facebook account that they have used through the same browser, Facebook can match their activity to their account [40].

2.5 Local storage

With the help of JavaScript, it is possible to store variables in key-value pairs straight in the browser, and retrieve them later when required [41]. This is made possible by the Web

Storage API and, although similar in function to cookies, much more straightforward an easier to use.

```
if (localStorage.clickcount) {
  localStorage.clickcount = Number(localStorage.clickcount) + 1;
} else {
  localStorage.clickcount = 1;
}
document.getElementById("result").innerHTML = "You have clicked the
button " +
localStorage.clickcount + " time(s).";
```

Figure 6. JavaScript code snippet that shows the use of local storage. The example counts the number of times a user has clicked a button. Example from W3Schools [42]**.**

Local storage can be used for tracking users the same way as cookies. A third-party can set a unique identifier in the browser's local storage that can be used to track them over multiple sessions [41]. The main advantages local storage has over cookies are the larger amount data it can store and the relative ease of use. Unlike cookies, data stored in local storage does not expire and it is kept indefinitely until the user or website deletes it.

Delfi uses local storage to store information required by the social login APIs and also some of the third-party tracking services included in Delfi use local storage. The following table describes the local storage variables set when visiting Delfi's website.

Table 1. Details on local storage variables set when visiting Delfi's website.

Variable	**Description**
loginApiAppleParams	Parameters for Apple's social login API.
loginApiFacebookParams	Parameters for Facebook's social login API.
loginApiGoogleParams	Parameters for Google's social login API.
loginApiParamsCache	Parameters for Delfi's own login API.
_cX_atfr	Unknown purpose. Naming indicates it belongs to the company Cxense.

_cX_expires	Unknown purpose. Naming indicates it belongs to the company Cxense.
_cX_G	Global ID cookie mapping different IDs together. Belongs to the company Cxense.
_cX_P	User session across sessions. Belongs to the company Cxense.
_hjid	A cookie set when a user first lands on a page. Used to persist the user ID on consecutive visits to the same site. Belongs to the company HotJar.

The variables belonging to HotJar and Cxense are copies of cookies with the same name and value. This is probably done due to redundancy and is an attempt to bypass privacy protections. If either local storage or third-party cookies are disabled, or cannot be used for some other reason, then the other one might still work.

3 Conclusions

The analysis performed in this work shows that Delfi itself does not seem to use any user tracking techniques or gather any unnecessary data on its users, only what is needed to create an Ekspress Meedia account if a user wishes to register one.

On the other hand, Delfi's authorized data processing partners are known to track users across different websites to create profiles on users that help serve targeted advertisements and have been implicated in privacy violations in the past. User data, such as visited pages, how much time a user spends on a ingle page, IP addresses, and technical details on the used device, is collected by companies such as Facebook, Google, Cxense, HotJar and Adform, which are using this data to improve targeted advertising and provide analytics solutions to Delfi.

The amount of third-party cookies and the number of analytics and advertisement partners Delfi uses is higher than on other popular Estonian news sites, but still on the lower end compared to other news sites in the European Union.

We found a privacy issue that results in Delfi currently not being GDPR compliant. Delfi fails to notify the user and does not ask for consent before setting third-party cookies on the user's browser. However, Delfi was contacted and they replied that they are aware of this issue and hope to have it fixed in the beginning of May.

To increase privacy while browsing Delfi, users can install privacy enhancing add-ons to their browser like DuckDuckGo Privacy Essentials[14] and choose to use a privacy conscious browser such as Firefox. These can help by disabling cookies set by known trackers and preventing other tracking techniques like tracking pixels and fingerprinting from being run in the user's browser.

[14] https://duckduckgo.com/app

4 References

[1] N. Confessore, “Cambridge Analytica and Facebook: The Scandal and the Fallout So Far,” 4 April 2018. [Online]. Available: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html.

[2] B. Wolford, “What is GDPR, the EU’s new data protection law?,” [Online]. Available: https://gdpr.eu/what-is-gdpr/.

[3] F. Manjoo, “Visited 47 Sites. Hundreds of Trackers Followed Me,” New York Times, 23 August 2019. [Online]. Available: https://www.nytimes.com/interactive/2019/08/23/opinion/data-internet-privacy-tracking.html.

[4] Gemius Estonia, “Gemius: Most visited websites in Estonia, March 2021,” 09 April 2021. [Online]. Available: https://www.gemius.ee/468/gemius-most-visited-websites-in-estonia-march-2021.html.

[5] A. Lerner, A. K. Simpson, T. Kohno and F. Roesner, “Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016,” in *Proceedings of the 25th USENIX Security Symposium*, Austin, 2016.

[6] B. Cyphers, “ Google’s FLoC Is a Terrible Idea,” 03 March 2021. [Online]. Available: https://www.eff.org/deeplinks/2021/03/googles-floc-terrible-idea.

[7] Privacy International, 23 October 2017. [Online]. Available: https://privacyinternational.org/explainer/56/what-privacy.

[8] Human Rights Careers, “Why Privacy Rights Are Important,” [Online]. Available: https://www.humanrightscareers.com/issues/reasons-why-privacy-rights-are-important/.

[9] Kantar Emor, “Privacy Rights and Personal Data Protection 2020 (from Estonian),” 2020.

[10] M. Murumaa-Mengel, P. Pruulmann-Vengerfeldt and K. Laas-Mikko, “The Right to Privacy as a Human Right and Everyday Technologies,” Estonian Institue of Human Rights, 2014.

[11] K. Litman-Navarro, “We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.,” [Online]. Available: https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html.

[12] B. Wolford, “Writing a GDPR-compliant privacy notice,” [Online]. Available: https://gdpr.eu/privacy-notice/.

[13] L. Olejnik, “Web browsing histories are private personal data - now what,” 24 August 2020. [Online]. Available: https://blog.lukaszolejnik.com/web-browsing-histories-are-private-personal-data-now-what/.

[14] Iubenda, “How to write a privacy policy,” [Online]. Available: https://www.iubenda.com/en/help/5525-cookies-gdpr-requirements.

[15] R. Koch, “Cookies, the GDPR, and the ePrivacy Directive,” [Online]. Available: https://gdpr.eu/cookies/.

[16] Ekspress Meedia AS, “Activites (from Estonian),” Eskpress Meedia AS, [Online]. Available: https://www.ekspressmeedia.ee/tegevusalad/veebimeedia/delfi/.

[17] Ekspress Meedia AS, “Privacy Policy,” [Online]. Available: https://www.ekspressmeedia.ee/partnerile/andmekaitse/privacy-policy/.

[18] Ekspress Meedia AS, “Cookie policy,” [Online]. Available: https://www.ekspressmeedia.ee/partnerile/andmekaitse/cookie-policy/.

[19] Privazyplan, “Recital 30,” [Online]. Available: https://www.privacy-regulation.eu/en/recital-30-GDPR.htm.

[20] CookieScan, “€39,000 fines for four companies breaking Cookie Law,” [Online]. Available: https://www.cookiescan.com/news/detail/fine-spanishSA-cookie-violations.

[21] CNIL, “Cookies: financial penalty of 35 million euros imposed on the company AMAZON EUROPE CORE,” 10 December 2020. [Online]. Available: https://www.cnil.fr/en/cookies-financial-penalty-35-million-euros-imposed-company-amazon-europe-core.

[22] CNIL, “Cookies: financial penalties of 60 million euros against the company GOOGLE LLC and of 40 million euros against the company GOOGLE IRELAND LIMITED,” 10 December 2020. [Online]. Available: https://www.cnil.fr/en/cookies-financial-penalties-60-million-euros-against-company-google-llc-and-40-million-euros-google-ireland.

[23] K. Koovit, “Fines for infringing data protection laws have been ridiculously low in Estonia (from Estonian),” 21 January 2021. [Online]. Available: https://majandus.postimees.ee/7161393/andmekaitsereeglite-rikkumise-eest-on-eestis-trahvitud-naeruvaarselt-vahe.

[24] O. Grauberg, “Estonia as one of the weakest links in the European data protection chain (from Estonian),” 02 January 2021. [Online]. Available: https://kpmglaw.ee/eesti_on_andmekaitses_euroopa_liidu_ueks_norgim_lueli.

[25] Intersoft Consulting, “Art. 15 GDPR Right of access by the data subject,” [Online]. Available: https://gdpr-info.eu/art-15-gdpr/.

[26] Intersoft Consulting, “Art. 28 GDPR Processor,” [Online]. Available: https://gdpr-info.eu/art-28-gdpr/.

[27] Google, “Secure your site with HTTPS,” [Online]. Available: https://developers.google.com/search/docs/advanced/security/https.

[28] A. Cahn, S. Alfeld, P. Barford and S. Muthukrishnan, “An Empirical Study of Web Cookies,” in *WWW '16: Proceedings of the 25th International Conference on World Wide Web*, 2016.

[29] Mozilla, “Using HTTP cookies,” [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies.

[30] D. Temkin, “Charting a course towards a more privacy-first web,” 03 March 2021. [Online]. Available: https://blog.google/products/ads-commerce/a-more-privacy-first-web/.

[31] StatCounter, “StatCounter Global Stats,” [Online]. Available: https://gs.statcounter.com/browser-market-share.

[32] D. Bohn, “Google to ‘phase out’ third-party cookies in Chrome, but not for two years,” 14 January 2020. [Online]. Available: https://www.theverge.com/2020/1/14/21064698/google-third-party-cookies-chrome-two-years-privacy-safari-firefox.

[33] T. Libert and R. K. Nielsen, “Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement,” May 2018. [Online]. Available: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-05/Third-Party%20Web%20Content%20on%20EU%20News%20Sites.pdf.

[34] IAPP, “Web Beacon,” [Online]. Available: https://iapp.org/resources/article/web-beacon/.

[35] B. Merchant, “How Email Open Tracking Quietly Took Over the Web,” 11 December 2017. [Online]. Available: https://www.wired.com/story/how-email-open-tracking-quietly-took-over-the-web/.

[36] H. Lauritzen, “Why do we need a cache-buster?,” AudienceReport, 26 December 2019. [Online]. Available: https://helpdesk.audiencereport.com/hc/en-us/articles/201421776-Why-do-we-need-a-cache-buster-.

[37] Facebook, “Facebook Pixel,” [Online]. Available: https://developers.facebook.com/docs/facebook-pixel.

[38] D. Nield, “Wired,” 12 January 2020. [Online]. Available: https://www.wired.com/story/ways-facebook-tracks-you-limit-it/.

[39] PassCamp, “Security and Privacy Risks when using Social Logins,” 17 January 2020. [Online]. Available: https://www.passcamp.com/blog/security-and-privacy-risks-social-logins/.

[40] A. S. John, “How Facebook Tracks You, Even When You're Not on Facebook,” 11 April 2018. [Online]. Available: https://www.consumerreports.org/privacy/how-facebook-tracks-you-even-when-youre-not-on-facebook/.

[41] W3C, “Web Storage (Second Edition),” 28 January 2021. [Online]. Available: https://www.w3.org/TR/webstorage/.

[42] W3Schools, “Window localStorage Property,” [Online]. Available: https://www.w3schools.com/jsref/prop_win_localstorage.asp.

[43] Adform, “Adform cookies,” [Online]. Available: https://site.adform.com/privacy-center/adform-cookies/.

[44] Mozilla, “HTTP headers,” [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/HTTP/Headers.

Appendix

The following table details all the cookies set on the user's computer when visiting Delfi. Descriptions of cookies are gathered from Web Cookies Scanner[15], CookieServe[16] or from documentation provided by the companies the cookies belong to. If a description is missing, then its due to no sufficiently reliable information being available. Parenthesis after the description contain the company associated with the cookie.

Table 2. Cookies set on the user's computer when visiting Delfi.

Domain	Name	Description
adform.net	C	Identifies if the user's browser accepts cookies or not (Adform).
clickonometrics.pl	CCMSESSID	Unique identifier for the current session (Clickonometrics).
hit.gemius.pl	Gdyn	Prevents the user from seeing a questionnaire they should not see or have already filled (Gemius).
hit.gemius.pl	Gtest	A tracking cookie. (Google)
doubleclick.net	IDE	Used to present relevant ads to the user (Google).
google.com	NID	Used to show relevant ads in Google services to signed out users. Also used for personalized autocomplete features in Google services (Google).
clickonometrics.pl	SERVERID	Assigns the user to a specific server to help with load balancing (Clickonometrics).
adform.net	TPC	Identifies if the user's browser accepts third-party cookies (Adform).
delfi.ee	_gfp_64b	Used to limit the number of times a user sees an advertisement (Google).
adsby.bidtheatre.com	_kuid	Cookie used to collect info on the user for the purpose of displaying targeted ads (BidTheatre).
s.delfi.ee	_edc	
s.delfi.ee	_edcCORS	

[15] https://webcookies.org

[16] https://www.cookieserve.com

delfi.ee	_edid	
delfi.ee	_edt	
delfi.ee	_fbp	A tracking cookie (Facebook).
delfi.ee	_ga	A cookie used visitor, session and campaign data. The cookie stores info anonymously and assigns a random number to each visitor (Google Analytics).
delfi.ee	_gat	Used to limit requests to the server (Google).
delfi.ee	_gat_rembi	
delfi.ee	_gid	Cookie used to collect anonymous analytics data (Goole Analytics).
delfi.ee	_hjAbsoluteSessionInProgress	Cookie used to detect the first pageview of a session (HotJar).
delfi.ee	_hjFirstSeen	Cookie used to detect if this is the user's first ever session (HotJar).
delfi.ee	_hjIncludedInSessionSample	Cookie used to determine if the current user is included in the daily data sampling (HotJar).
delfi.ee	_hjTLDTest	A cookie used to determine the most optimal generic cookie path to use (HotJar).
delfi.ee	_hjid	A cookie set when a user first lands on a page with the Hotjar script. Used to persist the user ID on consecutive visits to the same site (HotJar).
bidswitch.net	c	A cookie set by Magnite (formerly Rubicon Project). Purpose unknown.
delfi.ee	cX_G	Global ID cookie mapping different IDs together (Cxense).
delfi.ee	cX_P	User session across sessions (Cxense).
delfi.ee	cX_S	User session during a single session (Cxense).
delfi.ee	cX_T	A cookie used to find the top-level domain of the site (Cxense),
clickonometrics.pl	ccxid	
delfi.ee	cp_user_package_t	
delfi.ee	dcid	
delfi.ee	delfi-adid	

delfi.ee	evid_00XX	
facebook.com	fr	A cookie used to track logged out users across the web (Facebook).
cxsense.com	gckp	A cookie used to build a user profile across all sites on the Cxense network (Cxense).
de17a.com	guid	
de17a.com	guid2	
delfi.ee	test_cookie	A test cookie used by several providers to check if a browser accepts cookies.
bidswitch.net	tuuid	Indicates if the user has consented to the use of cookies or not (BidSwitch),
bidswitch.net	tuuid_lu	Contains a unique identifier that allows to track the user across multiple websites (Bidswitch).
adform.net	uid	Unique identifier used by Adform.
clickonometrics.pl	uint	
adnxs.com	uuid2	A tracking cookie used to track the user across multiple websites (Appnexus).

Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, Magnus Valgre

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   Tracking And Privacy: The Case of News Site Delfi

   supervised by Arnis Paršovs.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Magnus Valgre

07/05/2021