Lauri Pirilä

An overview on web cookies and privacy

Jyväskylän yliopisto
informaatioteknologian tiedekunta
2/21/2023

ABSTRACT

Pirilä, Lauri
An overview on web cookies and privacy
Jyväskylä: University of Jyväskylä, 2023.
Bachelor’s thesis, Information Systems Science
Supervisor: Kokko, Tuomas

This thesis researches web cookies and their use on the internet. This paper has been carried out as a literature review. The aim is to find out what cookies are, how they are used, what they are used for, how regulations affect cookies, and how are websites’ compliant with the regulations. The aim of this paper is also to provide the reader with useful information about the properties of cookies, their utilization on the server side, and their safety on behalf of information security and privacy. The results of this paper show what cookies are, what threats are associated with the use of web cookies, and how official regulations or laws affect the use and handling of cookies.

Keywords: Cookies, internet, website, HTTP (Hypertext Transfer Protocol), GDPR (General Data Protection Regulation), privacy, information security

TIIVISTELMÄ

Pirilä, Lauri
Yleiskatsaus evästeisiin ja yksityisyyteen
Jyväskylä: Jyväskylän yliopisto, 2023.
Tietojärjestelmätieteen kandidaatintutkielma
Ohjaaja: Kokko, Tuomas

Tämä tutkielma tutkii yleisellä tasolla evästeitä ja niiden käyttöä internetissä. Tutkielma on toteutettu kirjallisuuskatsauksena. Tavoitteena on selvittää, mitä evästeet ovat, mihin ja miten niitä käytetään, millä tapaa lainsäädännöt koskevat evästeiden käyttöä, sekä kuinka verkkosivut noudattavat lainsäädäntöä. Tavoitteena on myös välittää lukijalle hyödyllistä tietoa evästeiden ominaisuuksista, niiden hyödyntämisestä palvelinpuolella sekä niiden turvallisuudesta tietoturvan ja yksityisyyden kannalta. Tämän tutkielman tulokset näyttävät, mitä evästeet ovat, mitä uhkia tai haavoittuvuuksia evästeiden käyttöön liittyy, sekä miten viralliset säännökset tai lait vaikuttavat evästeiden käyttöön ja käsittelyyn.

Asiasanat: Evästeet, internet, verkkosivu, HTTP (Hypertext Transfer Protocol), GDPR (General Data Protection Regulation), yksityisyys, tietoturva

TABLE OF CONTENTS

1 Introduction

In the early days of the internet, a state management mechanism was needed for websites to remember session data for e-commerce applications, as well as for providing a more personalized user experience. One of the methods introduced to enable this was the use of HTTP cookies, which quickly became the standard for web browsing state management. Cookies are a small piece of data stored on the client computer by the server, and can be retrieved by the server on subsequent visits. (Kristol, 2001.) The privacy concerns regarding cookies were noted since the early days, and authorities have since set laws regulating the use of cookies. In the present day, websites' compliance with the data protection regulation laws seem to vary (Matte, Bielova, & Santos, 2020). When you enter a website for the first time, almost all of them ask for your consent in using cookies, in the form of a so-called "cookie banner". These banners don't necessarily give you a clear definition of what cookies are. Also, the reason why and how the website handles and utilizes cookies may often be unclearly defined. A privacy-concerned person might be cautious of consenting to cookies. People might want to know more about cookies, and how they should deal with them when they are asked about them during a website visit. This thesis, in the form of a systematic literature review, aims to answer the research questions set, thus providing the reader a better understanding of cookies' functionality, necessity and purposes, security concerns, and websites' methods of compliance with the regulations.

Most of the literature in this thesis has been searched from Google Scholar, with search words such as "cookies", "HTTP cookies", "cookie security", "GDPR" and "cookie privacy". Noteworthy sources of selected research literature are the IEEE Xplore Digital Library ACM Digital Library. The main research questions for this thesis are:

- What are HTTP cookies?
- How do cookies work?
- Are cookies a risk to online security and privacy, and if so, how?
- What kind of regulations are there regarding the use of cookies?
- How are websites' compliant with the regulations of cookies?

Chapter 2 explains the technical properties of cookies, their use, and their different categories. Chapter 3 presents security threats associated with the use of web cookies. Chapter 4 discusses cookie laws and regulations. Chapter 5 concludes the thesis with a summary of the information presented in this paper.

Findings of this thesis indicate that cookies are used for various purposes by several different parties, and that there are several security threats associated with cookies. Findings also show that GDPR and the ePrivacy Directive govern the appropriate handling of cookies, and adhering to these regulations demands several requirements from an organization or a web service.

2 AN OVERVIEW OF WEB COOKIES

This chapter gives an overview of web cookies – what is the history behind them, what are cookies, what different types of cookies exist, and why they are needed in modern web browsing. This paper uses terms “web cookie(s)”, “cookie(s)”, and “HTTP cookie(s)” interchangeably. Other types of cookies like zombie, flash, and edible cookies are not in the scope of the term “cookie” in this paper, unless specifically mentioned.

2.1 History

The term “cookie” was coined in the mid 90’s by Lou Montulli, an engineer at Netscape Communications Corporation. He was working on developing the Netscape Navigator web browser, and came up with the idea of using text files to store information about a user’s browsing session (Kristol, 2001). It is said that he named these files "cookies" as a reference to the "magic cookie" concept from UNIX computer science, where a small piece of data is used to remember a user's preferences or settings.

Cookies were first introduced publicly in the first version of netscape browser, in 1994 (Kristol, 2001). The first standard for cookies was introduced in the document RFC2109, in February 1997 (Kristol & Montulli, 1997; Cahn, Alfeld, Barford & Muthukrishnan, 2016). The aim of the introduction of cookies was to enable the web server and client to operate in a larger context, called a session (“session” not meaning a persistent network connection, but instead a logical session derived from cookies). Before that, the first early web browsers did not support state mechanisms, as each request and response was processed individually without any information being stored about previous or subsequent requests. (Kristol & Montulli, 1997.) Workarounds for the state mechanism before the introduction of cookies were, for example, embedding state information in a site’s URL address, or using the client's IP address to store state information. However, these methods were unreliable. URL-based state information would roll back to the previous state if the client clicks the browser's back-button, and it would increase network congestion. IP-based state information would also pose problems, for example a scenario where a website is used via a proxy, in which case all of the proxy’s users will appear as one user to the server. IP addresses can also be temporary, which means that state information will easily be lost if the client’s IP address changes. (Kristol, 2001.) Over the years, cookies have evolved and they have remained a central and essential part of the functionality of the internet.

2.2 What is a web cookie?

Web cookies are small amounts of data that a website stores on a user's device when they visit the website. They are used for storing browsing session data; to remember the user's preferences and activity on the site, such as shopping cart contents, login information, or language settings. (Kristol, 2001; Cahn et al., 2016.) The method of storage for cookies can vary depending on the web browser and operating system in use. Cookies are stored on every browser individually, and cookies stored in eg. Firefox are not visible to Google Chrome (Gaur, 2022).

Cookies are communicated between the browser and the server in the HTTP header field. During the initial interaction between a web browser and the server (shown in figure 1), the web browser first sends a HTTP GET request to the server, without sending a cookie (no cookie has yet been set to the browser, so no cookie is sent by the browser). The server responds by sending the cookie to the browser in the HTTP header field named "Set-Cookie", along with the HTTP message body. The web browser sends the obtained cookie to the server with subsequent requests within the header field named "Cookie". (Barth, 2011.)

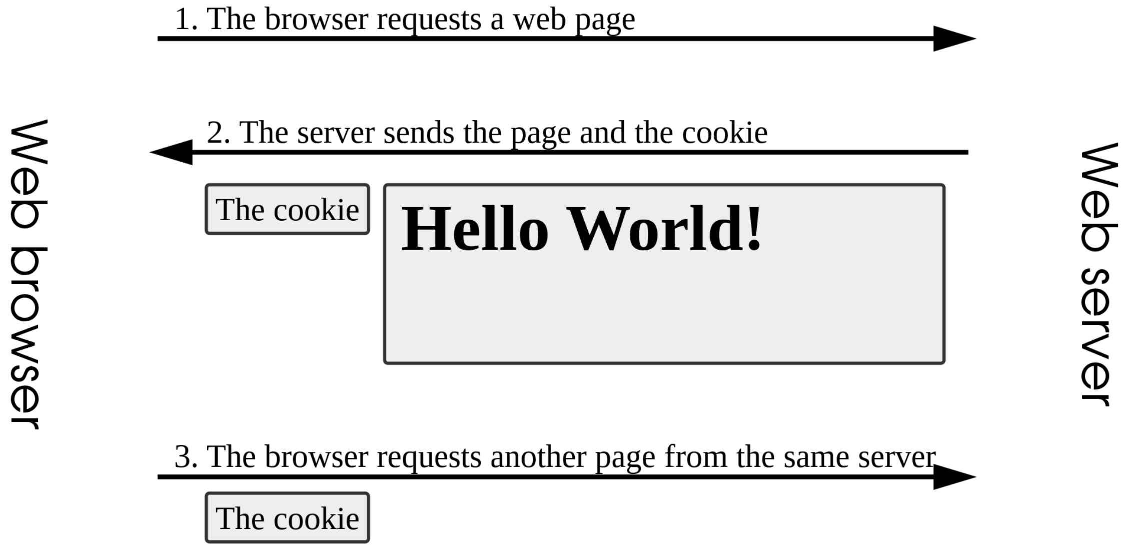

1. The browser requests a web page
Web browser
2. The server sends the page and the cookie
The cookie
Hello World!
Web server
3. The browser requests another page from the same server
The cookie

Figure 1 (Tizio, 2022.): A figure depicting the initial interaction between a web browser and a web server.

A simple cookie would look like the following (Cahn et al., 2016):

Name=Value; Host=example.com; Path=/account;
Expires=Tue, 1 Dec 2018 10:12:05 UTC; Secure;

The cookie consists of key-value pairs joined by an equals-sign ("="), in example: "*Host=example.com*", where the left side of the "="-sign specifies the name of the cookie, and the right side defines its value. Different key-value pairs are separated by a semicolon (";"). The cookie may also consist of different flags, which are not key-value pairs, but simple statements separated by a semicolon. In the above example, the "*Secure*" flag is set, which means that the cookie will only be sent via an encrypted connection (such as HTTPS). There are other types of cookie flags, such as the "HttpOnly" flag, which makes the cookie not obtainable by client-side scripts. (Barth, 2011.)

2.3 Cookie utilization

Cookies are used for essential web browsing activities. They allow the user agent to store items to a shopping cart, log in on a website, move around the website, and stay logged in. They can store the user's preferences on the website, such as layout settings or language. (Kristol, 2001; Cookiepedia, 2023a.) This way, they are very helpful to the users, unarguably providing them with a more pleasant browsing experience.

Cookies can be used to track a user's activity on the website. This type of tracking within the domain is called "web analytics". It makes it possible for the website host to improve the website by learning about, or emphasizing, the most popular ways of navigating the website. This can also enable the website to serve in a more targeted way, possibly providing the user with more worthwhile advertisements, content, or product suggestions. Although this type of tracking can be anonymous and doesn't require sensitive information, the EU's data protection laws still considers this to be private information. (Cahn et al., 2016; Bollinger, 2021.) Cookies can also be used to offer more personalized pricing on some services (Choe, King, & Matsushima, 2018).

Cookies might be used in a similar way to benefit other parties than the user browsing the website, or the website host. Cookies are used for targeted advertising and user profiling. Profiling or tracking web users could be valuable for many things, such as online advertisement purposes, or espionage. It is possible to track users across the web between different websites via third party cookies or similar cookie mechanisms. Data brokerage firms' and online advertisers' important goal is to amass as much information about a user in order to provide efficient targeted advertising (Cahn et al., 2016). It is worth noting that cookies are by no means the only way to track user behavior on the internet. Aside from cookies, a website is capable of identifying a user by, for example, an IP address. When combined with additional ambient information, it is possible to ensure that the activity carried by the browser represents a single user (Schwartz, 2001). Thus, there are other types of user tracking methods, such as fingerprinting, web beacons and super cookies (Cookiepedia, 2023b), which are not in the scope of this paper.

2.4 Different types of cookies

Cookies can be categorized to different types based on their properties or purposes. The following are some of the main different types of cookies:

Session cookies are cookies temporarily stored in the browser's memory. When the browser is closed, the cookies will be destroyed. They will remain even if the user temporarily navigates away from the website in the same browsing session. They can be used to store login credentials, in which case the user needs to log back into the website between different browsing sessions (browsing session ending when the user closes the browser). (Cookiepedia, 2023c; Barth, 2011.)

Persistent cookies are stored on the user's device or browser, and remain there even after the browsing session is over. Persistent cookies are set by the server with an expiry date. Users can delete the cookie before the expiry date ends. Websites can remove cookies from the user by sending the browser cookies with expiry date in the past, as when the client makes a request to the website with it, the cookie will be removed. Servers can update persistent cookies and move the expiry date further into the future. Like session cookies, they can also be used to remember a user's login information for a website, in which case the user doesn't necessarily need to log in to the website between different browsing sessions. (Barth, 2011; Cookiepedia, 2023c.)

First party cookies are cookies which are set by the website which the user has visited, or is currently on. They are used for essential web browsing purposes on a specific website. The host-attribute of a first party cookie is the same as the domain name on the address bar of the web browser. Only the host can set and retrieve the cookie. (Cahn et al., 2016.)

Third party cookies are cookies set by other sites than the site the user is visiting (a domain different from the one visible in the browser's address bar). 3rd party cookies are most commonly used for advertisement purposes, and they make it possible for advertisers to track users across multiple unrelated websites (Cookiepedia, 2023c). This is called "behavioral tracking" (Krishnamurthy & Wills, 2009), and it is one of the things at the heart of discussions of privacy concerns on the web. As third party cookies can link a single user across multiple visits on different websites (Hu & Sastry, 2019), third party cookie providers, such as Google Analytics have enormous power, since they can obtain a comprehensive view of the browsing history of a user. Browsers often have an option to block third party cookies. It does not necessarily mean that the user's information doesn't end up in the hands of third parties (Chen, Ilia, Polychronakis, & Kapravelos, 2021).

Secure cookies are cookies in which the secure-flag is set. This means that the cookie will only be transmitted via TLS and HTTPS. This helps against some of the security issues associated with cookies, such as eavesdropping. Unsecured HTTP communication is a common security threat with cookies. (Cookiepedia, 2023; Cahn et al., 2016; Sivakorn, Polakis, & Keromytis, 2016)

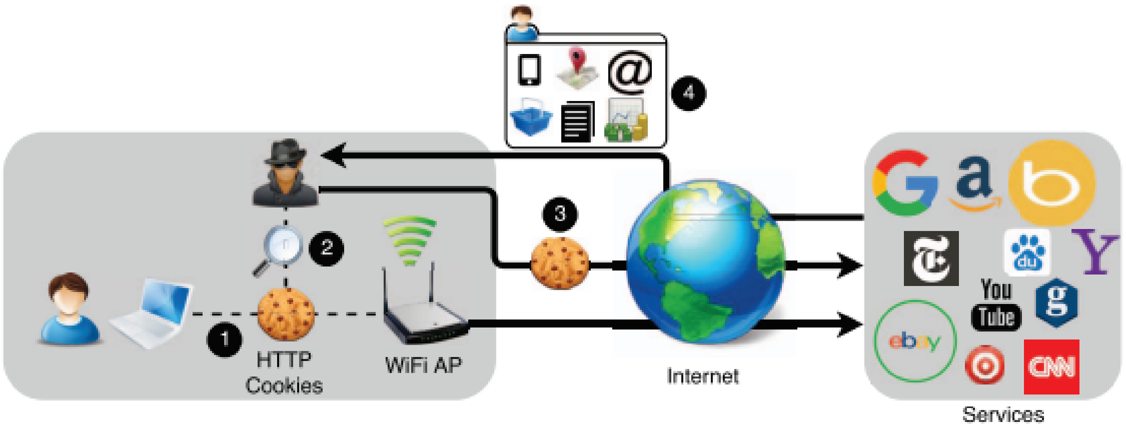
HttpOnly cookies are cookies in which the HttpOnly flag is set. When it is set, it means that the browser doesn't allow the cookie to be accessed on the client side by any scripts (ie. JavaScript). This protects the cookie from cross-site-scripting attacks, where a script would try to send the cookie's content to a third party. (Cookiepedia, 2023c; Cahn et al., 2016.)

3 The security of cookies

As cookies are sent in the HTTP header field, it is not possible to send any viruses or malware along with them directly. However, there are various information security threats associated with the use of cookies. If a cookie is not secured properly, using Transport Layer Security (TLS) and security-enhancing cookie flags, they are prone for different types of attacks. Numerous very popular websites, like Google, Bing, Yahoo and Amazon, have been, or are at risk of exposing personal information to hijackers, due to the lack of ubiquitous use of TLS (HTTPS). (Sivakorn et al., 2016.) Although some of the research on websites' security has been done before the widespread commandment of adhering to the GDPR principles, the methods of hijacking presented are still prevalent, as many (even popular) websites' compliance with the GDPR is still questionable, as shown by Matte et al. (Matte et al., 2020). Presented here are a few ways a user's web cookies could be obtained or exploited by an attacker or a third party.

3.1 Cookie hijacking

Cookie hijacking is a scenario where an attacker gains access to a person's cookies by monitoring their network traffic. Using a public wireless network, for example, in a coffee shop, makes it possible for a person to be vulnerable to cookie hijacking attacks. In cookie hijacking, the attacker gains access to the person's cookies when requests to a website are made in cleartext over an unencrypted connection. The website identifies the user by the cookies, so the attacker can gain access to the personalized version of the website made for the hijacked user, thus exposing the user's account and personal information to the attacker. (Sivakorn et al., 2016.)



A figure depicting a cookie hijacking attack. First, the user exposes the cookies to the attacker by making an unencrypted request to a website. An attacker can make a request using these cookies, and receive access to the user’s account and personal information. (Sivakorn et al., 2016.)

Physical presence, such as being in the same WiFi-network is not necessary in order to hijack a user's cookies if the attacker follows a more active approach. An attacker can inject content to the user’s browser in order for it to make requests to a vulnerable website and expose the cookies this way. Cookie hijacking is mostly reliant on insecure HTTP communication, so protective measures against it would be for the user and the server to communicate only via an encrypted channel by using TLS, more specifically HTTPS. It should be noted that it is important to enforce HTTPS in all stages of the communication, including the initial request by the browser, in order to prevent or reduce the possibility of cookies being hijacked by an adversary. (Sivakorn et al., 2016.)

3.2 Cross–site scripting (XSS)

If the user’s cookies do not have the HttpOnly flag set, they can be obtained by Cross-Site Scripting (also known as XSS). It is a type of web security vulnerability that allows an attacker to inject malicious scripts into a web page viewed by users. This can occur when a website doesn't properly validate user-supplied input and include it in the web page without proper escaping or sanitization. As a result, an attacker can inject client-side scripts such as JavaScript into the page, which can then be executed by other users visiting the page. With XSS, the attacker can steal sensitive information such as user passwords and session cookies, manipulate or deface the web page, and redirect users to malicious sites. An XSS attack can also be used to launch phishing attacks, spread malware, and conduct other types of cyber attacks. To prevent XSS attacks, it's important to properly validate and sanitize user-supplied input and filter or escape any special characters. A user should keep software and plugins up to date to address any security vulnerabilities. (Schiller, Fogie, DeRodeff, & Gregg, 2011.)

3.3 Cross-site request forgery (CSRF)

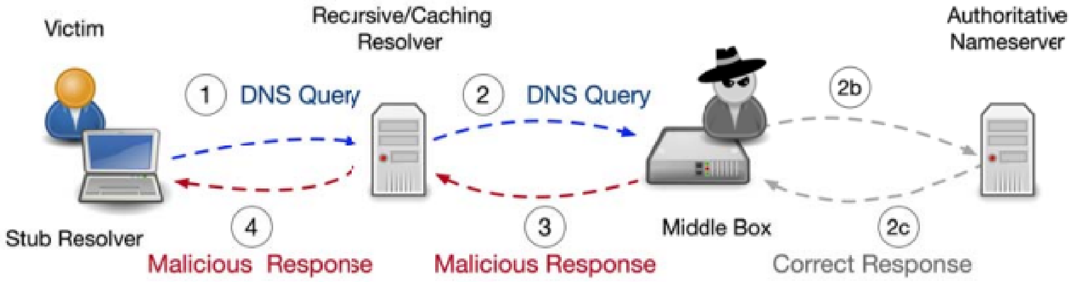
Cross-Site Request Forgery (also known as CSRF) is a security vulnerability that causes a victim’s browser to perform unintended actions on a website. This can be done by a hidden link or image, often disguised as a legitimate request to a website. When the link is clicked by the user, the website processes it as a legitimate request sent by the user. In other words, the attacker makes the victim submit a malicious request to the website, with the victim’s cookies included. This can lead to serious consequences for the victim, such as data theft, data manipulation or account takeover. To reduce the CSRF vulnerabilities, website developers should implement proper security measures, such as security tokens, two-factor authentication for sensitive requests, as well as sanitization, filtering and validation of user-supplied input. (Lin, Zavarsky, Ruhl, & Lindskog, 2009.)

3.4 Browser hijacking

Installing virus or malware on a computer can have serious consequences. Browser hijackers are malware that modify internet browser settings. This type of malware can also contain keyloggers, which monitor keystrokes, capturing any sensitive information the victim might enter into websites. The modified browser can redirect the victim to malicious websites when using it, as well as steal the browser's cookies. To avoid being the target of a browser hijacking attack, one should at least have an antivirus installed and functional, have an up-to-date browser, and they should not install any suspicious free software. (Malwarebytes, 2023.)

3.5 DNS hijacking

Domain Name System (DNS) is a central part of the web infrastructure, and allows websites to be recognized by their domain names (eg. www.example.com) instead of an IP address. DNS hijacking is an attack that can lead to cookies and personal information being captured by an attacker.. In DNS hijacking, DNS queries by the victim are compromised. The DNS queries can be directed to a compromised, or an attacker-controlled DNS server. This can be done by the attacker eg. via malware that is installed on the computer, by accessing and modifying the user's router settings, by compromising a nameserver, or by a "man in the middle" (MITM) type of attack. Different types and methods of DNS hijacking makes defending against it not that straightforward. (Houser, Hao, Li, Liu, Cotton, & Wang, 2021.)



A figure depicting a DNS MITM attack. (Houser et al., 2021)

3.6 CNAME cloaking

The DNS records are controlled by the domain owner, and the domain owner can map a (sub)domain to an IP address or another domain name, which the DNS will look up via a Canonical Name (CNAME) record. This can be useful, as media can be provided from a content delivery network on behalf of the first-party server, without security suspicions. (Ren, Wittman, De Carli, & Davidson, 2021.)

CNAME cloaking refers to a technique that allows third parties to present themselves as first parties, thus eluding the same-origin policy, which states that cookies will only be sent to the first party (the cookie's host). Scripts from the CNAME domain have permissions that can be contrary to the website administrator's intentions, such as giving access to first party cookies. This method can make it possible for advertisers and trackers to evade ad-blockers. Serious consequences such as account takeover can result from authentication cookies' exfiltration. However, authentication cookies alone may not be enough for an attacker to impersonate a user. (Ren et al., 2021.) Mitigation against CNAME cloaking as a user can be difficult. Ad-blockers such as the UBlock Origin do have blocklists that prevent sensitive information from being leaked to known third parties, but these blocklists are manually curated, so there is no absolute guarantee of protection. (Ren et al., 2021.)

3.7 Rotten cookie

Kwon et al. showed that even though HTTPS is used in between client-server communication, there are still potential vulnerabilities in cookies. In a cookie-cutter attack, even if TLS is used, miscommunications between TLS and HTTP allow the attacker to remove secure–enhancing cookie flags by closing the connection. TLS or HTTPS itself lacks a security mechanism to protect cookies. TLS still has many vulnerabilities, although the newest specification aims to eliminate some of them. Some servers might still be liable to the vulnerabilities as they ensure backwards compatibility. In a *rotten cookie* attack, once the attacker has found a vulnerable server and target client, they carry out a Man-in-the-Middle (MITM) attack by altering TLS fragments. Their goal is to steal private cookies from the victim by removing the flags associated with the cookies. The root cause of a rotten cookie attack is the reuse of nonces in AES-CGM. Generating a nonce in a secure way would be an effective mitigation in the web transport layer. This attack method is rare and not that significant, as it is not usual to reuse nonces. Only a handful of websites are susceptible to this kind of attack. (Kwon, Nam, Lee, Hahn, & Hur, 2020.)

3.8 Other security threats

There are numerous ways to identify a user online, whether it be by using cookies or other online tracking methods, such as fingerprinting. Like previously mentioned, third party tracking and user identification without consent can be considered a security and a privacy issue.

Browsers support blocking third party cookies, and some do it already by default, or are planning to do so in the near future. However, blocking third party cookies does not itself guarantee that cookies and private information will not end up in the hands of third parties via other methods, such as external cookies set up by JavaScript code that abuse first-party cookies (Chen et al., 2021).

Even using a tor-browser does not necessarily guarantee anonymity, as it has been demonstrated that unencrypted HTTP traffic is flowing through tor exit nodes, exposing vast amounts of private user information (Sivakorn et al., 2016). Also, it is shown that private information about cryptocurrency transactions can be leaked to trackers. Online trackers are able to see details of payment flows sufficiently enough to link a purchase uniquely to a transaction on the Bitcoin blockchain. Some merchants might additionally leak private information (such as names and email addresses) to trackers, allowing trackers to link user transactions to the user's real identity as well as the user's web profile. (Goldfeder, Kalodner, Reisman, & Narayanan, 2017.)

4 Cookie regulations

The issues that have required official regulations on cookies are associated with things such as privacy, profiling and tracking. A large portion of the economy behind the internet is on the behalf of advertising. Internet advertising is a huge and fast growing industry, and it was estimated to be worth around 227 billion dollars in 2018. (Sanchez-Rola, Dell'Amico, Kotzias, Balzarotti, Bilge, Vervier, & Santos, 2019.) Total digital ad spending was shown to be over 600 billion USD in 2021, by a Statista report. Digital advertising is projected to grow exponentially to over 870 billion USD by the year 2026. (Geyser, 2022.)

The digital advertising industry heavily relies on personalizing advertisements through targeted ads, which increase their appeal to users. However, the collection of data required for personalization also poses significant privacy concerns. Web cookies, in particular, can allow advertisers to access a significant portion of a user's browsing history, potentially revealing sensitive information such as medical conditions and political opinions. (Sanchez-Rola et al., 2019.)

Privacy concerns regarding the use of cookies have been pointed out and discussed over the years ever since the introduction of cookies (Schwartz, 2001; Kristol, 2001; Krishnamurthy & Wills, 2009). Privacy is considered as freedom from unauthorized surveillance and it is a human right according to the United Nations (Dabrowski, Merzdovnik, Ullrich, Sendera & Weippl, 2019) and European Union (Degeling, utz, Lentzsch, Hosseini, Schaub, & Holz, 2019).

3.1 GDPR and e-Privacy Directive

The General Data Protection Regulation (GDPR) was enacted on May 25, 2018 in the European Union (EU) with the aim of safeguarding the online privacy of its citizens (Hu & Sastry, 2019). It is a law concerning the use of personal data. The General Data Protection Regulation of the European Union has had a notable effect on the internet. It replaced some of the previous directives and became immediately enforceable in all EU states. It defined strict and enforceable legal requirements on service providers' as well as third parties' personal data collection and tracking. It mandated that service providers and third parties comply with its regulations within a two-year time frame. (Bollinger, 2021.)

Previous privacy directives such as the ePrivacy Directive (ePD) and the Directive 95/46/EC have been set as an attempt to improve web users' privacy by defining guidelines and requirements that businesses and web hosts should comply with when handling users' personal data. (Bollinger, 2021.) It is worth noting that The ePrivacy Directive supplements the GDPR with additional regulations regarding the handling of personal data in the electronic communication sector, including websites. The GDPR is a regulation that is directly enforceable in all European countries, while the ePrivacy Directive is a directive that each member state must implement within its own national laws. (Matte et al., 2020.)

Since its implementation, end-users have become aware of the ePrivacy Directive due to the widespread presence of a "Cookie Bar" on most websites. This bar informs users of the existence of tracking mechanisms and seeks their consent for their usage. The reach of the ePrivacy Directive extends beyond Europe as any web service with users in the EU must comply with its regulations (Trevisan, Stefano, Bassi & Marco, 2019.), and the same goes for GDPR. Nevertheless, previous directives before the GDPR have frequently failed to achieve their intended purpose due to inconsistent implementation across EU member states, hindering the legislation's ability to effectively influence the actions of third parties (Bollinger, 2021).

Cookies can contain personal data. The principles regarding the processing of personal data, as stated in the GDPR Article 5 (EUR-Lex, 2016), are the following (cited precisely, word-to word):

Personal data shall be;

a. *processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');*
b. *collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');*
c. *adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');*
d. *accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');*
e. *kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');*
f. *processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').*

…

(end of citation)

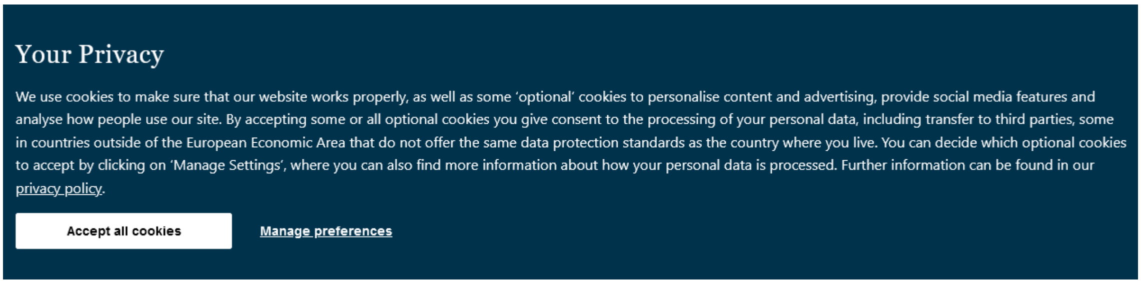
Thus, the GDPR demands organizations to implement appropriate technical and organizational measures to ensure the protection of personal data collected through the use of cookies. This includes measures to ensure the confidentiality, integrity, availability and removability of the data. (EUR-Lex, 2016.)

Article 6 of the GDPR states that a data controller needs a legal basis to process identifiable information. This is generally either willful consent of the user, or one of the exceptions mentioned, such as the processing being necessary for the purpose the person is subject to. This necessary information can be eg. a pizza delivery service storing a customer’s home address in order to be able to deliver a pizza. (Dabrowski et al., 2019; EUR-Lex, 2016; Hoofnagle, Sloot, & Borgesius, 2019.) GDPR governs the processing of all kinds of personal data, but more specific rules for the handling of web cookies can be found in the ePrivacy Directive. Moreover, the European commission has published a proposal for an ePrivacy Regulation, which aims to replace the ePrivacy Directive. (Hoofnagle et al., 2019; Borgesius, van Hoboken, Fahy, Irion, & Rozendaal, 2017.)

The failure to comply with the GDPR principles can result in a notable enforcement fine. Multi-million dollar fines have been set for failure to comply with the regulations. (Ruohonen & Hjerppe, 2022.) Overall, the GDPR has required organizations to be very aware and conformed to the data protection principles. Many organizations have, for example, had data protection awareness seminars for majority, if not all, personnel within the organization in order to make the workforce more knowledgeable of the principles of data protection. (Sirur, Nurse, & Webb, 2018.)

3.2 Websites’ compliance with GDPR and ePD regarding cookies

The use of tracking cookies and similar tracking technologies for behavioral advertising requires the recipient's prior consent. Consent notices are not a new phenomenon, as design recommendations for them date back to 2001. Consent notices became common when ePD came into effect in 2009. Consent notices have many forms, such as banners, prompts and overlays. Some of these provide options to manage the use of cookies, and some do not provide other options than to accept the contents with an “OK” button. (Bollinger, 2021.) “Nudging” users to consent to cookies is common behavior, and it is shown that privacy-aware users need to often spend more time configuring their cookie settings (Hils, Woods, & Böhme, 2020).



Your Privacy

We use cookies to make sure that our website works properly, as well as some 'optional' cookies to personalise content and advertising, provide social media features and analyse how people use our site. By accepting some or all optional cookies you give consent to the processing of your personal data, including transfer to third parties, some in countries outside of the European Economic Area that do not offer the same data protection standards as the country where you live. You can decide which optional cookies to accept by clicking on 'Manage Settings', where you can also find more information about how your personal data is processed. Further information can be found in our privacy policy.

Cookie banner of SpringerLink. (https://link.springer.com/)

As GDPR came into effect, some existing consent notices needed upgrading. Complying with the requirement for "specific consent" entails categorizing cookies based on individual usage. "Informed consent" means that each cookie category had to be provided with detailed description. "Explicit consent" mandates that the consent for data collection cannot be assumed implicitly. The process is further complicated by the need to communicate information about the user's consent with several third parties, which significantly increases the cost and effort of implementing compliant solutions. (Bollinger, 2021.)

The Interactive Advertising Bureau (IAB) Europe produced the Transparency and Consent Framework (TCF) to aid tracking and advertisement industries in the consent collection. They also introduced the notion of Consent Management Providers (CMPs). (Matte et al., 2020.) Consent management has been widely outsourced to the CMPs. They emerged in the years following the GDPR to standardize consent collection on the web. CMPs create an ecosystem of consent by defining legal terms and conditions, presenting them to users through an embedded consent dialogue, storing the resulting signal, and sharing it with third-party vendors. It thereby involves users, websites, and third-party vendors. (Hils et al., 2020.)

Still, several studies have shown that websites' compliance with the GDPR and cookie regulations is not pervasive. Hu & Sastry showed in 2019 that non EU-sites are less likely to offer options to manage their privacy preferences, as well that the availability of different cookie consent options varies with website category (Hu & Sastry, 2019). Matte et al. studied over 1400 websites with TCF consent banners, and showed that among these, 236 websites nudge users to accept consent with pre-selected options and 27 websites don't pay regard to the users choices at all, allowing consent even if the user has opted out. Over half of the 560 websites that were in closer inspection had at least one violation. (Matte et al., 2020.)

IAB's TCF and CMPs seek to help organizations and services to adhere to the ePD and GDPR, and makes it easier for organizations to manage consent without reading and fathoming the entire official regulation documents thoroughly. The use of TCF and CMPs doesn't explicitly mean that a website ubiquitously follows the laws and regulations. (Matte et al., 2020.)

5 Conclusions

The main focus of this research was to find out what cookies are, and what kind of security or privacy concerns are related to them. The research questions of this thesis were:

- What are HTTP cookies?
- How do cookies work?
- Are cookies a risk to online security and privacy, and if so, how?
- What kind of regulations are there regarding the use of cookies?
- How are websites' compliant with the regulations of cookies?

The above questions have been answered throughout this paper. This thesis investigated various aspects of HTTP cookies, such as their appearance, nature, purposes, security, privacy, and lawfulness. Cookies are an important part of the everyday use of the internet. As such, they have become essential in the use of state management, session management, targeted advertising, ease of use, and other types of information.

Cookies are set to the web browser by the web server. The browser sends the cookies back to the server if host-associated cookies are stored within the browser. Cookies can mainly be divided into first party cookies, which are supposed to be sent only to the host who set the cookies to the browser, and third party cookies, which are sent to other parties than the website the user is visiting. (Barth, 2011.) Third party cookies are at the center of privacy issues on the internet. Blocking third party cookies is not necessarily sufficient to prevent cookies from being transferred to third parties by other means. (Chen et al. 2021.)

There are several information security vulnerabilities associated with cookies, and malicious attacks targeting cookies differ by their purpose, methods, and severity. To ensure cookie safety, it is good practice to use a secure, up-to-date browser, possibly with extensions that enhance safety. The ubiquitous use of Transport Layer Security, more specifically HTTPS, prevents many sorts of cookie vulnerabilities. (Sivakorn et al., 2016; Lin et al., 2009.) Modern browsers typically have the option to enable HTTPS-only setting, preventing HTTP from being used to expose personal information via unencrypted communication.

Improving the security of internet use is a continuous process, with new regulations and methodology being introduced continually to better enhance communication-, server-, and user safety. Cookies can contain sensitive personal information. Due to privacy concerns, laws and regulations have been set to protect web users. The most notable of these is the GDPR, which affects an enormous number of individuals, services and organizations. (EUR-Lex, 2016; Hoofnagle et al., 2019)

This paper covers cookies on an overall level. More technical details can be found in other research literature which address more specific aspects of cookies. There exists extensive amounts of research on cookies, which cover subjects such as cookie security vulnerabilities, use of cookies for tracking purposes and regulation abidance. These areas are still a good field for future research on cookies, for obtaining up-to-date information. A good future research area could also be measuring how cookie banners’ appearance and functionality affects how people consent to cookies, and how they react to consenting.

Bibliography

Barth, A. (2011). *RFC6265.* Internet Engineering Task Force (IETF) https://www.rfc-editor.org/rfc/rfc6265

Bollinger, D. (2021). *Analyzing Cookies Compliance with the GDPR* [Master Thesis, ETH Zurich]. https://doi.org/10.3929/ethz-b-000477333

By Tizio - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=1577024

Cahn, A., Alfeld, S., Barford, P., & Muthukrishnan, S. (2016). An Empirical Study of Web Cookies. *Proceedings of the 25th International Conference on World Wide Web*, 891–901. https://doi.org/10.1145/2872427.2882991

Chen, Q., Ilia, P., Polychronakis, M., & Kapravelos, A. (2021). Cookie Swap Party: Abusing First-Party Cookies for Web Tracking. *Proceedings of the Web Conference 2021*, 2117–2129. https://doi.org/10.1145/3442381.3449837

Choe, C., King, S., & Matsushima, N. (2018). Pricing with Cookies: Behavior-Based Price Discrimination and Spatial Competition. *Management Science*, *64*(12), 5669–5687. https://doi.org/10.1287/mnsc.2017.2873

Cookiepedia. (2023a). *All about cookies.*
https://cookiepedia.co.uk/all-about-cookies.

Referenced 1/10/2023.

Cookiepedia. (2023b). *Other tracking technologies.* *https://cookiepedia.co.uk/Content/OtherTrackingTechnologies*. Referenced 1/10/2023.

Cookiepedia. (2023c). *Types of cookies. https://cookiepedia.co.uk/types-of-cookies*. Referenced 1/10/2023.

Dabrowski, A., Merzdovnik, G., Ullrich, J., Sendera, G., & Weippl, E. (2019). Measuring Cookies and Web Privacy in a Post-GDPR World. In D. Choffnes & M. Barcellos (Eds.), *Passive and Active Measurement* (pp. 258–270). Springer International Publishing.

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., & Holz, T. (2018). We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. *CoRR*, *abs/1808.05096*. http://arxiv.org/abs/1808.05096

EUR-Lex. *Document 02016R0679-20160504* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504

Gaur, Chandan. (2022). *Local Storage vs. Session Storage vs Cookies.* Xenonstack. https://www.xenonstack.com/insights/local-vs-session-storage-vs-cookie. Referenced 2/21/2023.

Geyser, W. (2022). Internet Advertising Statistics - Rise of Mobile and Ad Blocking. Influencer Marketing Hub.

https://influencermarketinghub.com/internet-advertising-statistics/. Referenced 2/12/2023.

Goldfeder, S., Kalodner, H. A., Reisman, D., & Narayanan, A. (2017). When the cookie meets the blockchain: Privacy risks of web payments via cryptocurrencies. *CoRR*, *abs/1708.04748*. http://arxiv.org/abs/1708.04748

Hils, M., Woods, D. W., & Böhme, R. (2020). Measuring the Emergence of Consent Management on the Web. *Proceedings of the ACM Internet Measurement Conference*, 317–332. https://doi.org/10.1145/3419394.3423647

Hoofnagle, C. J., Sloot, B. van der, & Borgesius, F. Z. (2019). The European Union general data protection regulation: What it is and what it means. *Information & Communications Technology Law*, *28*(1), 65–98. https://doi.org/10.1080/13600834.2019.1573501

Houser, R., Hao, S., Li, Z., Liu, D., Cotton, C., & Wang, H. (2021). A Comprehensive Measurement-based Investigation of DNS Hijacking. *2021 40th International Symposium on Reliable Distributed Systems (SRDS)*, 210–221. https://doi.org/10.1109/SRDS53918.2021.00029

Hu, X., & Sastry, N. (2019). Characterising Third Party Cookie Usage in the EU after GDPR. *Proceedings of the 10th ACM Conference on Web Science*, 137–141. https://doi.org/10.1145/3292522.3326039

Krishnamurthy, B., & Wills, C. (2009). Privacy Diffusion on the Web: A Longitudinal Perspective. *Proceedings of the 18th International Conference on World Wide Web*, 541–550. https://doi.org/10.1145/1526709.1526782

Kristol, D., Montulli, L. (1997). *RFC 2109.* Network Working Group. https://www.rfc-editor.org/rfc/rfc2109

Kristol, D. M. (2001). HTTP Cookies: Standards, Privacy, and Politics. *ACM Trans. Internet Technol.*, *1*(2), 151–198. https://doi.org/10.1145/502152.502153

Kwon, H., Nam, H., Lee, S., Hahn, C., & Hur, J. (2020). (In-)Security of Cookies in HTTPS: Cookie Theft by Removing Cookie Flags. *IEEE Transactions on Information Forensics and Security*, *15*, 1204–1215. https://doi.org/10.1109/TIFS.2019.2938416

Lin, X., Zavarsky, P., Ruhl, R., & Lindskog, D. (2009). Threat Modeling for CSRF Attacks. *2009 International Conference on Computational Science and Engineering*, *3*, 486–491. https://doi.org/10.1109/CSE.2009.372

Malwarebytes. (2023). Browser Hijacker. .*https://www.malwarebytes.com/blog/threats/browser-hijacker*. Referenced 2/14/2023.

Matte, C., Bielova, N., & Santos, C. (2020). Do Cookie Banners Respect my Choice? : Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. *2020 IEEE Symposium on Security and Privacy (SP)*, 791–809. https://doi.org/10.1109/SP40000.2020.00076

Ren, T., Wittman, A., De Carli, L., & Davidson, D. (2021). An analysis of first-party cookie exfiltration due to cname redirections. *Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb)*.

Ruohonen, J., & Hjerppe, K. (2022). The GDPR enforcement fines at glance. *Information Systems*, *106*, 101876. https://doi.org/10.1016/j.is.2021.101876

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A., & Santos, I. (2019). Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, 340–351. https://doi.org/10.1145/3321705.3329806

Schiller, C., Fogie, S., DeRodeff, C., & Gregg, M. (2011). *Infosecurity 2008 threat analysis*. p. 65-205. Elsevier.

Schwartz, J. *Giving the Web a Memory Cost Its Users*

Privacy, NY Times. September 2001. https://www.nytimes.com/2001/09/04/business/giving-web-a-memory-cost-its-users-privacy.html. Referenced 2/21/2023.

Sirur, S., Nurse, J. R. C., & Webb, H. (2018). Are We There Yet? Understanding the Challenges Faced in Complying with the General Data Protection Regulation (GDPR). *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*, 88–95. https://doi.org/10.1145/3267357.3267368

Sivakorn, S., Polakis, I., & Keromytis, A. D. (2016). The Cracked Cookie Jar: HTTP Cookie Hijacking and the Exposure of Private Information. *2016 IEEE Symposium on Security and Privacy (SP)*, 724–742. https://doi.org/10.1109/SP.2016.49

Trevisan, M., Stefano, T., Bassi, E., Marco, M., & others. (2019). 4 years of EU cookie law: Results and lessons learned. *Proceedings on Privacy Enhancing Technologies*, *2019*(2), 126–145.

Zuiderveen Borgesius, Frederik and van Hoboken, Joris V. J. and Fahy, Ronan P. and Irion, Kristina and Rozendaal, Max, An Assessment of the Commission's Proposal on Privacy and Electronic Communications (June 7, 2017). Directorate-General for Internal Policies, Policy Department C: Citizen's Rights and Constitutional Affairs, ISBN: 978-92-846-1100-3, Available at SSRN: https://ssrn.com/abstract=2982290