GDPR - Required Changes By Website Operators And Technical Ways To Check For Compliance

Maximilian Muth
Advisor: Quirin Scheitle
Seminar Future Internet SS2018
Chair of Network Architectures and Services
Departments of Informatics, Technical University of Munich
Email: muthm@in.tum.de

ABSTRACT

Tracking technology on websites has become more advanced and easy to deploy. While it allows website operators to gain useful insights about their site visitors and even additional business models by using the collected data, it reduces the privacy of the consumers.
In 2011, the European Union tried to improve this situation by introducing a directive which required website owners to ask the visitor for consent to store cookies in the web browser and to give site visitors the right to refuse the use of cookies. The directive's goal was to increase awareness about which data is collected and to improve the consumers online privacy. In 2016 the European Union approved the General Data Protection Regulation (GDPR) which aims at protecting online privacy of consumers in multiple ways.
In this paper we present the most important changes by GDPR for website owners and highlight what changed since the "cookie law" directive from 2011. We will also take a look at a website and analyze whether it is already GDPR compliant or what they would need to change to be compliant when the GDPR is implemented on 25th May 2018.
One of our findings from that analysis is, that although a small company already put in effort in order to be compliant, they apparently still did some mistakes.
As a final reminder, this paper is no legal advise but an opinion from an engineering point of view.

Keywords

GDPR; General Data Protection Regulation; European Union Data Protection; GDPR Compliance

1. INTRODUCTION

The protection of natural persons by protecting their personal data is a fundamental right according to Article 8(1) of the Charter of Fundamental Rights of the European Union [7]. One of the recitals in the preamble of the General Data Protection Regulation (GDPR)[1] is the following, which clearly phrases the privacy problems with today's online tracking technologies:

> *Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet Protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags.*
> *This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.*[7]

A simple JavaScript snippet by a tracking or analytics service provider is enough to track website visitors over a period of time, get detailed insights about the users interaction or even recordings of the screen the user sees with precise mouse movement tracking.
By using tracking cookies, it is even possible for providers to track a certain user across multiple websites. Another issue is that advertisement and tracking providers share identifiers and data between each other which allows them to create more detailed user profiles.

This results in increased privacy concerns in many cases and often the site visitors are not even aware of the fact that they are tracked and profiled in such detailed ways by the website operators.
Sometimes it might even not be on purpose that providers implement tracking possibilities in their products, like Apple did with their Apple Push Notification Service (APNs) which enabled user tracking based on TLS Client Certificate Authentication [13].

The European Union tried to improve this situation by introducing a EU directive in 2011[2] which required website owners to ask the visitor for consent to store cookies in the web browser and to give site visitors the right to refuse the use of cookies. The directive's goal was to increase awareness about which data is collected, to reduce the amount of stored data to the necessary needed data and to improve the consumers online privacy.
To enhance the regulations on data protection, the parliament of the European Union approved the GDPR on 14 April 2016 after four years of debate. The General Data Protection Regulation is a replacement for the Data Protection Directive 95/46/EC which has been in place since 1995.

The GDPR will be implemented - and thus enforced - by 25

[1] Official identifier: REGULATION (EU) 2016/679

[2] `http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm`

doi: 10.2313/NET-2018-11-1_08

May 2018 and there will be penalties on organizations which do not comply to it by that date.

In the first part of this paper we will discuss the new aspects and rule sets of the GDPR with a focus on the technical implementation.

In the second part we will summarize what website operators need to change in order to be compliant to the GDPR and what possibilities exist for consumers to check whether websites are compliant.

2. OVERVIEW

The GDPR introduces some new key concepts which we discuss in the following chapter. We also show under which conditions the GDPR applies.

2.1 New Topics And Regulations

In this section we pick a couple of interesting key changes or strengthened topics in the GDPR. We're focusing on topics which are related to the rights of consumers towards website operators.

2.1.1 Lawfulness of Processing

One of the basic concepts of the GDPR is that the processing of personal data is prohibited unless stated otherwise by the law. The most important reasons why a processing might be allowed are:

1. Processing is necessary for the provided service or contract (Art. 6.1.b GDPR[3])
2. *"Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject"* (Art. 6.1.f).
3. The user provided consent (Art. 6.1.a)

Details about the consent are discussed in the upcoming section. The first reason however, is interesting since it seems to not really be clear when a site operator's legitimate interests (German: *Berechtigte Interessen)* legitimate him to process data without the user's consent.
It looks like for example online marketing and analysis are such legitimate interests [6]. At the first impression it seems like *legitimate interests* might be a loophole for providers to process data without the need of the user's consent. However, in recital 47 GDPR it says the following:

> *"At any rate the existence of a legitimate interest would need careful assessment including* ***whether a data subject can reasonably expect*** *at the time and in the context of the collection of the personal data* ***that processing for that purpose may take place****".*

So while the juristical community discusses what exactly legitimate interests are in certain cases [6], in our opinion it looks like the GDPR tries to be strict regarding the term of legitimate interests.
At the current point in time the discussion seems to be still ongoing and we should await first judgments in order to better understand this topic.

2.1.2 Consent

Some of the most interesting changes are related to the consent of the *data subject* (the site visitor) about how the subject's data is processed and stored.
Some key elements of GDPR related to consent are that the request for consent must be obviously presented as such a request and the phrasing must be easy to understand and no legalese (Art. 7.2). It also must be clear why the requested data is needed and what the site operator is using it for (Art. 7.). After consent was given, it shall be as easy to withdraw the consent as it is to give (Art. 7.3).
This is interesting because some websites seem to hide the reasons on why they ask for certain personal data in large privacy policy documents. However, they now need to make sure to tell the user in an understandable way what the personal information is used for at the moment the user gives its consent.
Furthermore the user needs to give consent for storing personal data before the data is actually collected and stored. This is important for the usage of third-party tools like analytic or advertisement frameworks which by default might set cookies and collect personal data like the used browser, resolution, IP address etc. before the user gave his consent. Processing of personal data without the consent of the user or other reasons is not legal (Art. 6.1) and it is the website operator's duty to ask for that consent before collecting personal data.
Site operators need to make sure that they can prove to supervisory authorities that a given user gave consent for the processing of his personal data (Art. 6.1).

2.1.3 Information and Access to Personal Data

Users have a right to request information about whether specific information are stored and processed by a site operator (Art. 15.1.e), as well as for what purpose the information is used and whether it is used for automated decision-making programs or profiling (Art. 15.1.a,e).
But most of all, the user has a right to request the personal information that the operator stores about him. The operator must hand out all stored data in an electronical format, without charging a fee. Website operators seem not to need to provide an automated process for sending out the requested data, but can handle those requests for data by hand, which might be important for small website operators.

2.1.4 Right to Erasure

Right to erasure - also called "Right to be Forgotten" - enables consumers to request the deletion of all personal data stored and generated by the website operator. One of multiple conditions must be met in order to be able to lawfully request the removal from a *controller* (like a website operator), most importantly: the data is not used anymore for the purpose for which it was collected (Art. 17.1.a) or the user has withdrawn his consent (Art. 17.1.b).

[3] Used law text citation style: *Art. 7.2.a* refers to point (a) of the second paragraph of Article 7. If not stated otherwise, the cited lawtext is the GDPR

So whenever a user wishes to have his data removed he can withdraw his consent and request the deletion.
Another point which enhances the user's privacy is that the controller needs to forward the request for erasure to other parties to whom the personal data was disclosed (Art. 19.). However, the law text also states that this is only needed *unless this proves impossible or involves disproportionate effort* (Art. 19.), so it looks like controllers might argue to not do this due to high effort implementing such a solution.

2.1.5 Right to Object

According to Art. 21.1.a a user can restrict what processing his personal data is used for. This is especially relevant in the case of automated profiling. If the personal data is used for direct marketing purposes, the user can object that usage at any given time (Art. 21.2).
In reality it is questionable how well this works - it is possible that site operators simply stop providing their service to the user in case he requests to restrict the processing of his personal data. As a result, users might not request restriction because they want to use the given service.
At this point in time, we didn't find any resources to validate this claim.

2.1.6 Data Portability

The right for data portability extends the right for access to personal data (Art. 15.) in the way that the user is allowed to take the personal data and give it to another controller / website operator. The first operator is not allowed to hinder the usage of the data by another controller in any means (Art. 20.1).
This idea seems great since it aims at impeding the vendor lock-in strategy [8] which tries to make users stay by making moving to another provider complicated.
The law allows users to migrate between different service providers and even makes sure that the exported data is *"in a structured, commonly used and machine-readable format"* (Art. 20.1).
Nonetheless it is questionable how well the exported data can be used for migrating to another service, since the domains where a *"commonly used format"* like vCard for contact information[4] or iCal for calendar definitions exists, is rather limited in our opinion. The use of common (but non domain-specific) data formats like JSON, XML or CSV is highly beneficial nevertheless.
Phrased differently: What is a *commonly used format* of the data in a social network? Still, other service providers or open-source communities might come up with automated import routines based on this law.

2.1.7 Breach Notification

If a personal data breach occurs, the controller must notify the supervisory authority of the breach within 72 hours (Art. 33.1). A supervisory authority is determined by the member states themselves. In Germany the federal states are responsible in providing such an authority and in Bavaria the supervisory authority is *Bayerisches Landesamt für Datenschutzaufsicht* (BayLDA)[5].
In Germany, data breaches already must be reported according to paragraph 42a BDSG[6] in case the personal data is highly sensitive (like banking or health information) and as a second condition that there is a high risk of big impairments for the subject [1]. So while in Germany a law already required notifications to supervisory authorities, it was pretty loose and many incidents would not require a notification.
This might be different in the future: With the GDPR, every personal data breach needs to be reported to the supervisory authority, which is an improvement of the previous regulation. However, it is not required to notify the affected users unless it *is likely to result in a high risk to the rights and freedoms of natural persons* (Art. 34.1).
In such a case the controller needs to notify the user without any delay and also provide details about the breach like categories and subjects of the affected data as well as consequences (Art. 34.2). The notification must be plain text and easy to understand.
It is important to note, that the supervisory authority has the ability to require the controller to notify the data subject if the controller did not do so by himself.
An interesting exception where a notification is not needed is when the personal data is encrypted (Art. 34.3.f). As a result, the data subject might not be informed that encrypted personal data was leaked. We could argue that this is an issue since the attacker might not be able to break the encryption of the data on the date of the breach. But maybe in the future it is feasible.

At the time of writing, information are appearing that claim the company Cambridge Analytica collected personal information of multiple million Facebook users in order to target US electors with highly purposeful and targeted Facebook advertisements before the presidential election in 2016.[11]
Facebook is being criticized as well, for example because they did not notify the affected users that a third party illegally collected their personal data [11]. The discussion is still going on and it is too early to draw conclusions about it in this paper.

This current event stresses the importance of clearer and stricter rules and supervisory related to user's personal data.

2.2 Area Of Application

In the following section we discuss under which conditions the GDPR applies, especially under which conditions a website operator needs to make sure to comply and in which cases consumers have the rights GDPR gives them like the above mentioned strictly regulated request for consent or access to personal information.

When analyzing whether a given law applies in a certain context we need to take a look at two areas of application: material scope (German: *Sachlicher Anwendungsbereich*) and territorial scope (German: *Räumlicher Anwendungsbereich*).

2.2.1 Material Scope

Material scope is given when (semi-)automated processing of personal information occurs or when the processing is done manually but the data is part of a *filing system* as described

[4] `https://tools.ietf.org/html/rfc6350`

[5] `https://www.lda.bayern.de`

[6] Bundesdatenschutzgesetz

in Art. 2.1.
Furthermore material scope is not fullfilled in some defined cases, most importantly it does not apply when a natural subject processes his own personal information (Art. 2.2.c).
According to Art. 2.2.a the GDPR also does not apply when the processing of personal data occurs *"in the course of an activity which falls outside the scope of Union law"*. We're not quite sure how it is determined whether an activity is outside the Union law's scope or not, but we could imagine it is related to the territorial scope we discuss in 2.2.2.

We also see, that the material scope is only fulfilled if the processed data is *personal* (German: *personenbezogen*). The European Union provides a definition of personal data in Art. 4. as follows:

> *"'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;"*

This definition is rather coarse and does not provide any examples. However, it also states that personal data must not only be a single data item like a name, but that the combination of non-personal information might also be *personal data*.

Let's think about the power of combining data with a small example: Person P provides the following information to website operator A: City, gender, age group. To another service provider B P gives information about his profession, the range of his income, as well as his first name. While it is hard to track P down with each of those information separately, it is a lot more feasible by combining the two data sets. Of course, for combining, a common identifier is needed, which might be *P's* IP address, or a third-party cookie both websites store and read.
Combining data sets from multiple websites is helpful for creating user-profiles which allows operators to have more information available that can be used for eg. making online advertisement more efficient. So it is good to see the definition also covering the combination of multiple data sets.

2.2.2 Territorial Scope

The law only applies if the material scope and the territorial scope are both fulfilled. According to Art. 3.1 territorial scope is fulfilled if the context of the establishment, controller or processor is in the European Union. It is important to note that it is not relevant whether the processing itself takes place in the Union.
The regulation also applies if the controller or processor is not located in the Union, but the data subject (eg. site user) is, under the conditions that the goods or services are provided in the EU (Art. 3.2.a).
To compare this to the previous situation, territorial scope did not change for organizations located in the EU. Nonetheless, GDPR broadens the territorial scope so that EU privacy rules also apply for organizations outisde the Union if they operate in the Union (like selling their goods or services) - a situation which is referred to as Extraterritoriality[12].
As a result, non-EU organizations might be required to comply to the GDPR for their customer segment within the EU.

Besides privately held companies, it is also interesting to take a look at foreign governments. Some countries are implementing their own privacy endangering laws, like for example the recently discussed *CLOUD Act* in the US[2]. This bill would allow US law enforcement to force companies to hand out personal data about US citizens without the need of search warrants nevertheless whether the data is stored in the US or outside.
However, GDPR explains in Art. 48. that such a disclosure of data is only possible if there are international agreements which explicitly allow it. As a result, it might be possible that US companies must not need hand out data about US citizens if it is stored in the EU. Note, that at the current time, this is speculative and we'd need to wait for further juridical discussions or court judgments.

3. NEEDED CHANGES FOR OPERATORS AND CONTROL POSSIBILITIES

In the previous sections we outlined the most important changes of the GDPR in a theoretical context. Now, we want to focus on website operators and take a deeper look at what they need to change in order to be compliant with the GDPR.
Furthermore, we want to investigate for which of those changes a user can control whether the site operator is compliant or no and which of those changes could be possibly checked automatically.

3.1 Needed Changes

3.1.1 Google Analytics

Google Analytics[7] (GA) or Matomo (previously Piwik)[8] are wide-spread third-party integrations for tracking and analyzing user behavior on websites.
These frameworks usually store a wide range of metrics, like IP addresses, operating system, browser, browser settings, language, viewport resolution, HTTP header information like referrers and much more. Furthermore each page visit as well as the visit duration and interactions within the page are stored.
The question is, what of these information are personal information so the GDPR applies (see. 2.2). According to a judgment in 2016 of the European Court of Justice, IP addresses are personally identifiable information[9], so the site operator needs to ask for consent before allowing his analysis framework to collect and store the IP address.
In this section we will focus on Google Analytics because it has the largest market share. Anyways, most points are valid for other tools like Matomo as well.
There are two possible solutions for website operators to deal with IP addresses in GA:

[7] https://analytics.google.com/
[8] https://matomo.org/
[9] Judgment in legal matter C-582/14, European Court of Justice

1. Anonymize IP address, so it is not personally identifiable anymore [6]
2. Ask for consent before initializing GA

While the second approach is a bit more difficult to implement (the site needs to remember the state of the consent and the operator cannot just simply put the GA JavaScript snippet in the website), the first approach can be easily achieved. However, both approaches require the operator to take action because GA does not anonymize IP addresses by default.
If IP anonymization is enabled, GA will set the last 8 bit of the IP address to zero (or the last 80 bit for IPv6 addresses).

There are four ways to achieve IP anonymization in GA:[9]

- Set the *anonymizeIp* option when initializing GA via JavaScript: `ga('set', 'anonymizeIp', true)`. This globally enables IP anonymization for every event GA tracks on the site.
- Set the *anonymizeIp* flag for single events: `ga('send', 'pageview', 'anonymizeIp': true);`
- Set the *anonymizeIp* flag via Google Tag Manager (*gtag.js*)
- Set the *anonymizeIp* flag via the web GUI of Google Tag Manager

Although

3.1.2 Webserver Logs

As argued in the previous section, IP addresses are personal information. It is questionable whether monitoring of the webserver access log is a legitimate reason to store the IP address without the user's consent.
The request for consent is mostly on the client-site after the webpage has loaded, however, the webserver already logged the request at that point of time. As a result it is strongly recommended to anonymize the IP addresses in the webserver's log.
Furthermore, depending on the website, personal information might also be part of the URL path and an operator should anonymize this as well if he didn't clearly got the consent of the user.

Since writing log filters and anonymizing scripts is time consuming, it might be a good alternative to disable webserver access logs completely if the operator is not actively monitoring and analyzing them.

3.1.3 Newsletter

Many websites provide email newsletters to their users. Users need to sign up for the mailing list in order to receive the website operator's content per mail.
As it has already been before the GDPR, sending marketing mails to users without their consent is not legal. GDPR strengthens the way controllers need to verify that the user gave his consent, so according to Art. 7.1 the controller must be able to demonstrate that a given user gave consent to receive marketing emails.
It is noteworthy to repeat that this request for consent must have been clear and obvious to the user. Pre-ticked checkboxes for example are not legal (Recital 32, GDPR), the same goes for automatic newsletter signups for example after an order in an online shop.
Also, it is not allowed to require a sign-up for the newsletter in order to receive a service like the download of an eBook (Art. 7.4). This topic seems to be heavily discussed within the juristic community since it partly renders freemium models[10] useless.[5]

Since the controller must be able to demonstrate the consent, a *soft opt-in* is not enough. Soft opt-in means that a user gave his consent to receive the newsletter on the website itself by ticking a checkbox. Since anyone can do so with any email address, it is required to do an *double opt-in* which basically is a mail asking for the consent after the requested to be put on the newsletter list on the webpage.[5]

3.1.4 Social Media Integrations, Comment Systems, Gravatar, Ad Networks

Many website operators integrate some third-party tools like social media integrations (Facebook, Twitter, Xing, Linkedin, ...), comment systems like Disqus, avatar systems like Gravatar or advertisment networks like AdSense or Doubleclick. The challenge is, that many of those integrations send the user's personal data like IP addresses to the service provider [10].

For each of those integrated third-party tools, the website operator is in charge of requesting consent for the forwarding of personal data to the third-party provider. So the operator should mention all those services and state what the data is used for when requesting consent of the user. If the user did not yet provide consent, those third-party integrations should not be initialized.

3.1.5 Privacy Policy

The privacy policy should clearly state which data the website operator collects and stores, as well as which data is provided to third-party integrations including the reasons.
If the operator is relying on the user's consent to do so, the request for consent should be repeated in the privacy policy.
For some third-party integrations like Goolge Analytics detailed information need to be provided, like for example a way to opt-out of the data collection (remember, we argued that it is GDPR compliant to use GA with an anonymized IP without asking for consent, so the operator needs to make sure to provide a way to allow the user to prevent the pseudoanonymized data collection of GA).
In the case of GA there are two ways to achieve an opt-out, one of them is by setting a cookie and the other option is to suggest the user to install a browser extension provided by Google[11][6].

3.2 Technical Ways To Check For Compliance

One of the problems with many privacy related laws and regulations is that it is often hard to check for the users

[10] `https://en.wikipedia.org/wiki/Freemium`

[11] `https://chrome.google.com/webstore/detail/google-analytics-opt-out/fllaojicojecljbmefodhfapmkghcbnh`

whether controllers really behave accordingly. This is especially the case for "average" or non-technically skilled users. In this chapter we discuss which of the above mentioned measures a website operator should take can be controlled and checked by users - both, in a manual and maybe even automated way. By finding possibilities to automatically check for conformity and making the test results publicly available, it is imaginable to help users to preserve their privacy.

3.2.1 IP Anonymization Of Google Analytics

One of the most important measures a user can check is, whether Google Analytics anonymizes IP addresses. Because if not - the user would have needed to provide consent. In 3.1.1 we investigated the different ways how website operators can enable the IP anonymization in GA.

A user can manually check, whether IP address anonymization is enabled or not, by inspecting the source-code of the website and searching for the JavaScript snippet which sets the anonymization flag (`ga('set', 'anonymizeIp', true);`). If this flag is present, the page is probably anonymizing the user's IP address.

If the flag is not set in the body of the page, or the user want's to make sure that the flag is not overridden by single events[12], he can check the network communication between his browser and Google Analytics servers. This can be easily done with the Developer Tools in Google Chrome or Firefox.
Each event which is tracked in GA will perform a *HTTP GET* or *POST* request to *https://www.google-analytics.com/r/collect*[13] with additional data being passed as URL parameters or in the HTTP POST body. If IP anonymization is enabled (either by setting the global flag or by configuring it on the given event), the request contains the parameter `aip=1`, for example: *HTTP GET https://www.google-analytics.com/r/collect?v=1aip=1&...*. If the *aip* parameter is not present, the given event will be stored without an anonymized IP address.[9]

The monitoring of the outgoing AJAX requests might be even checked automatically in order to detect whether a given page is not compliant to the GDPR. One solution might be a browser extension[4] which analyzes the requests to the servers of GA and alerts the user in case the IP address is not anonymized. Such a browser extension could also store this information for later research.
On a larger scale it should be possible to run a web crawler (with JavaScript enabled) and test this on an arbitrary number of pages and log the results to a database for further analysis.

3.2.2 Newsletter Double-Opt-In

As elaborated in 3.1.3 a double-opt-in is required for newsletter sign-ups. This means, if a user subscribes to an e-mail newsletter on a website, he must receive a mail asking for confirmation before the operator is allowed to send him mails via the newsletter.

Obviously, this can be easily checked manually. Furthermore, it would be possible to implement automatic checks for the double-opt-in by using a web-crawler which performs sign-up events on newsletters and a program which is connected to the same e-mail account in order to check whether a confirmation mail has arrived.
There are some challenges, one would face when implementing this system. For example it is not given that the domain of the website where the crawler signed up for the newsletter is the domain-name of the mail server from which the confirmation mail arrives. A possible solution to this might be using a mail server which supports mail aliases and sign-up with a uniquely identifiable e-mail address for each website.

3.2.3 Privacy Policy

If a user wants to check the privacy policy, the first check might be whether the policy says which data is collected and for what reason. Furthermore, he needs to make sure that third-party integrations like Google Analytics or Facebook are mentioned if they are embeded in the website (the user might take a look in the source-code of the given page or use a browser extension like Wappalyzer[14] to detect those third-party integrations).
For integrations like GA an opt-out possibility must be proposed.

Besides, it might be hard for the user to evaluate whether the privacy policy conforms to all given laws, since it is not trivial to for a layman.

A check which can be automatically done is to search for keywords like *Google Analytics* or *Matomo* within the privacy policy if the crawler detects that such third party integrations are used (by either looking for corresponding JavaScript objects or cookies).

3.2.4 Breach Notification

In order to check whether the controller notifies a user about a data breach, the user needs to know that a breach occurred. This, however, is a circular condition, so it is probably non-trivial to check whether operators notify accordingly.
An approach which might be working was introduced by Joe DeBlasio et al. with a system called Tripwire[3] which is based on password-reuse. The idea is to use unique e-mail addresses to register on websites and to use the same password for the website and the email account itself. You then monitor over a longer period of time whether a successful login to the email account occurs. If such a login happens, it is interpreted as compromise of the the website for which the email address was used because the attacker used the password of the website account to access the corresponding mail account (the attacker checked whether the victim re-used his password).

While this might work in our situation, it is obviously not a perfect test environment for checking breach notification

[12] Each manual event which is sent to GA could override the anonymizeIp flag if the developer wants it to behave like this. However, this is probably a rare case. An event can be any interaction like a page view or a click on a button.

[13] In case of the legacy *ga.js* the target is `http://www.google-analytics.com/_utm.gif`

[14] `https://github.com/AliasIO/Wappalyzer`

compliance to the GDPR.
First of all, it only works if the passwords get stolen or forwarded to third-parties. However, there might be other personal data that leaks which requires a notification.
Next, although some websites will do so, most websites probably don't store passwords in plaintext but hashed using salted cryptographic hash functions like PBKDF2 or bcrypt, thus, an attacker might not be able to retrieve the actual password from the leaked hashed password in a short amount of time. As a result, it is not really possible to check the time it takes for the controller to notify the data subject.

Beside checking the controller, we could also check how the supervisory authority handles information of breaches by site operators. This way we could understand how well authorities in different countries implement their part of the GDPR and react to breach notifications by controllers.
Therefore, a fake website with fake users could be set up. We could then fake a breach and notify the corresponding supervisory authority of the breach and log how they react.

While this paper focuses on websites, it is noteworthy that data-stealing and breach notifications is also a highly interesting topic for mobile applications.

3.2.5 Access To Personal Data & Data Portability

It might be interesting to check whether a service sends out all information it stores about a user when the user requests access to its data or whether the exported data set is not complete. Therefore, a user could manually add personal data to a given service, exactly logging which information were provided. The user can then compare his log with the data the operator sends him after a request according to Art. 15..

The right for access to personal data can also be used to find out what other data sources a controller used in order to complement a user profile.

3.2.6 Measures Which Are Hard To Check

As we discussed it is feasible to check some measures manually or even automatically. However, there are a couple of measures which are harder to check or where the possible approaches are not feasible in reality:

1. Does the operator store records of processing activities? (Art. 30.)
2. Do webserver logs contain personal data?
3. Is personal information given to third-party providers?
4. Is personal data really deleted after an according request? (Art. 17.)
5. Usage of personal data for additional/other purposes than consent was given for? (Art. 7.)

4. WEBSITE ANALYSIS FOR COMPLIANCE

For this section, we pick a website and analyze it for compliance with the GDPR. We therefore use the techniques described in 3.2 and also focus on the mentioned measures, namely IP Anonymization, Newsletter and Privacy Policy. At the current point in time - where the GDPR is not implemented yet - it is hard to check for other measures like Access To Personal Data or Breach Notifications.

4.1 PlusPeter

PlusPeter is a brand provided by *PlusPeter GmbH, Berlin* which offers online printing services for students.
We decided to investigate PlusPeter because it is a rather small start-up and they collect personal information about students which makes them attractive to analyze since they are collecting personal data which the users want to know well protected.

4.1.1 Overview And Objects Under Analyis

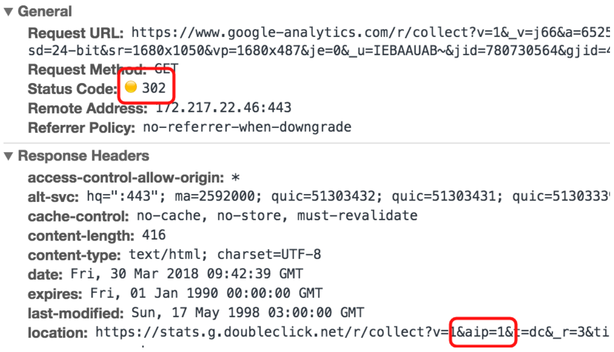
There are two domains which belong to the service. The first domain is `www.pluspeter.com` which is running a WordPress instance and provides landing pages and detailed information about their services.
The second domain is `app.pluspeter.com` which runs the web application where students can login and use the service.
We are taking a look at both domains.

4.1.2 www.pluspeter.com

Third Party Integrations

The site uses Google Analytics which is integrated via Google TagManager. Furthermore, one of Google's advertisment networks - Doubleclick - is used.

When inspecting the network traffic as in Figure 1, we first see a request to *google-analytics.com/r/collect* which indicates that an event is being tracked in GA.
This request does not contain the *aip* parameter and thus, does not instruct GA to anonymize the IP. However, the GA service returns a status code of *302* and redirects the request to the same endpoint on the domain *stats.g.doubleclick.net*, which the browser will fetch in a second request.
This second request to the doubleclick network contains the *aip* parameter, and thus, instructs GA to anonymize the IP address (3.2.1).



```
▼ General
    Request URL: https://www.google-analytics.com/r/collect?v=1&_v=j66&a=6525
    sd=24-bit&sr=1680x1050&vp=1680x487&je=0&_u=IEBAAUAB~&jid=780730564&gjid=4
    Request Method: GET
    Status Code: ● 302
    Remote Address: 172.217.22.46:443
    Referrer Policy: no-referrer-when-downgrade
▼ Response Headers
    access-control-allow-origin: *
    alt-svc: hq=":443"; ma=2592000; quic=51303432; quic=51303431; quic=5130333
    cache-control: no-cache, no-store, must-revalidate
    content-length: 416
    content-type: text/html; charset=UTF-8
    date: Fri, 30 Mar 2018 09:42:39 GMT
    expires: Fri, 01 Jan 1990 00:00:00 GMT
    last-modified: Sun, 17 May 1998 03:00:00 GMT
    location: https://stats.g.doubleclick.net/r/collect?v=1&aip=1&t=dc&_r=3&ti
```

Figure 1: Network request to GA on pluspeter.com

We performed this check on multiple subpages and with different interactions within each page.

To sum it up, the first request to GA did not provide the flag for IP anonymization, but according to the HTTP response code, the request was not processed by the GA collector but just redirected to a domain belonging to Doubleclick. This second request uses IP anonymization, so we can say that www.pluspeter.com uses GA with anonymized IP addresses.

Newsletter

Although the privacy policy mentions a newsletter, PlusPeter does not seem to provide a public newsletter we could check for compliance. However, we can inspect the rest of the privacy policy.

Privacy Policy

Google Analytics is correctly mentioned and the privacy policy[15] (last change 30.03.2017) also states how to disable the tracking of GA. The policy also covers the data which is collected when using the contact form of the site.

The business model of PlusPeter probably includes using the provided data about students (for example the topic of studies and the current semester) to allow companies to add targeted advertisements to the documents PlusPeter prints for the students for free. This is also mentioned in the privacy policy, where the policy explains that personal data is used for self-promotion of own services, as well as the promotion of services of other companies.
The document also says that personal data is forwarded to selected companies for postal advertisements. This is noteworthy, since it allows PlusPeter to not only use the data to offer targeted advertisments to their customers, but it also allows them to forward the corresponding data as is to companies.
Note, that according to the GDPR the users are able to object the forwarding of their personal data (2.1.5).

Access To & Removal Of Personal Data

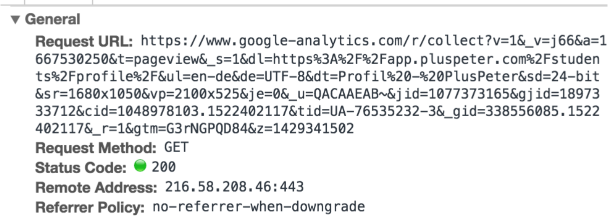
PlusPeter's privacy policy already contains a section related to the access to personal data where they state that the user can request insights about what data PlusPeter stores. The user can also request the data to be changed or removed. These sections have probably been preparations for the implementation of GDPR in May 2018.

The public website (`www.pluspeter.com`) does not have any further forms and seems not to collect any other data as already discussed.

4.1.3 *app.pluspeter.com*

Third Party Integrations

The app also uses Google Analytics as we can see by inspecting the network traffic of the browser.



▼ General
Request URL: https://www.google-analytics.com/r/collect?v=1&_v=j66&a=1667530250&t=pageview&_s=1&dl=https%3A%2F%2Fapp.pluspeter.com%2Fstudents%2Fprofile%2F&ul=en-de&de=UTF-8&dt=Profil%20-%20PlusPeter&sd=24-bit&sr=1680x1050&vp=2100x525&je=0&_u=QACAAEAB~&jid=1077373165&gjid=1897333712&cid=1048978103.1522402117&tid=UA-76535232-3&_gid=338556085.1522402117&_r=1&gtm=G3rNGPQD84&z=1429341502
Request Method: GET
Status Code: ● 200
Remote Address: 216.58.208.46:443
Referrer Policy: no-referrer-when-downgrade

Figure 2: Request to GA on app.pluspeter.com

Figure 2 shows that the tracking using GA within the app does not anonymize the IP address - unlike `www.pluspeter.com`.

It is noteworthy that this is not only the case for logged in users, but also if new users visit the registration page[16].

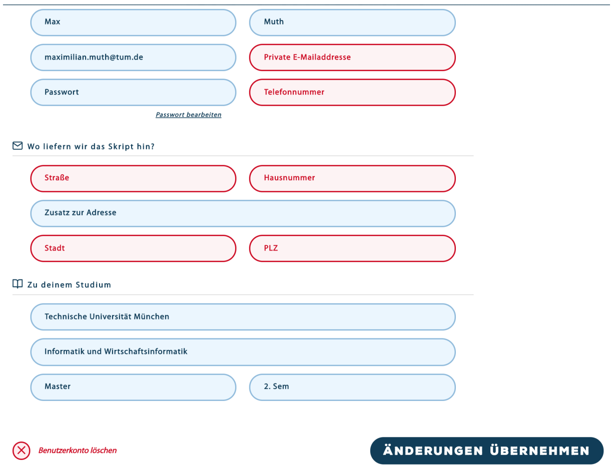
In this case, we did not provide consent for collecting our IP address on the registration page. As a result, this is probably not compliant with the GDPR (s. 2.1.2).

Request Of Consent

According to Art. 7.7 GDPR the website operator needs to clearly state what data is collected and for which purpose it is used. It is not enough to simply state it in legalese in an enormous privacy policy.

In the case of PlusPeter, there are two forms where the user activly enters data.
The first form is the registration form in which the user types in his email address and a password.
The second form is located in the profile settings where the user enters his personal data as well as information about the study progress (name, birthday, university- and private mail address, phone number, address of residence, university, degree, major / study path and the current semester of enrollment).

As a screenshot of the second form shows (figure 3), there is no text stating what the entered data is used for, nor what exactly is stored.
This is probably not compliant to the GDPR (s. 2.1.2).



Max
Muth
maximilian.muth@tum.de
Private E-Mailaddresse
Passwort
Telefonnummer
Passwort bearbeiten

Wo liefern wir das Skript hin?
Straße
Hausnummer
Zusatz zur Adresse
Stadt
PLZ

Zu deinem Studium
Technische Universität München
Informatik und Wirtschaftsinformatik
Master
2. Sem

Figure 3: Profile Information on app.pluspeter.com

Privacy Policy

The privacy policy we discussed for `www.pluspeter.com` is also applicable to the application at `app.pluspeter.com`.

In the section about GA the policy states, that the IP addresses will be anonymized. This is true in case of `www.pluspeter.com`, but is not true on `app.pluspeter.com` as shown above (even for visitors who are not registered for the service).

[15] `https://web.archive.org/web/20180330092452/https://www.pluspeter.com/datenschutzerklarung/`

[16] `https://app.pluspeter.com/students/registration/`

doi: 10.2313/NET-2018-11-1_08

As a result, this claim in the privacy policy is wrong. We are not sure how penalties for misinformation in privacy policies are regulated, it would probably need a lawyer to look at the specific case in order to name a penalty.

Summary
While it is interesting to see, that the privacy policy already covers some changes related to the GDPR with the date of last change on March 2017 (more than one year before the implementation of the GDPR), PlusPeter's websites still have some behaviors which are likely not compliant with the GDPR.

The app does not anonymize IP addresses when sending tracking events to GA, although the privacy policy states so and the visitors did not gave consent for that.

Furthermore, the app doesn't state what the collected data is used for next to the form. The information is partly available in the privacy policy, but that's not enough for being compliant with the GDPR (2.1.2).

Nevertheless, we see that PlusPeter put in some effort in trying to be compliant, like anonymizing IP address collection on the WordPress site or stating the possibility to request personal data. Regardless of these efforts, they still made some mistakes.

5. CONCLUSION

We presented the new key concepts of the GDPR which will be implemented May 25th, 2018. The focus was identifying needed changes website operators need to perform in order to be compliant with the new ruleset.
Most important topics we discussed were the general lawfulness of processing, the way it is required to get consent from users, concepts like access to personal data and its removal and how to handle data breaches.
With these understandings, we identified the required changes for most website operators, namely not collecting personal data without the user's consent with forms, integrations like Google Analytics or webserver logs, changing the way newsletter sign-ups are performed and monitored (Double-Opt-In), changes in the integration of social media services and comment systems as well as required changes in the privacy policy.

While discussing all of this, we noticed that it might be hard for smaller website operators to understand which changes are needed and what behavior is not allowed in order to be compliant. This was also shown in our analysis of the websites by PlusPeter GmbH, who already put in some effort to be compliant with the GDPR, but still missed some points or did mistakes.

Overall, it looks like the GDPR aims at improving the user's privacy by putting the user more in control of what data is processed in which ways. On the other hand, it increases the workload for small site operators to be compliant which might be even not viable for some operators.

As a final reminder, this paper is no legal advise but an opinion from an engineering point of view.

6. REFERENCES

[1] Bayerisches Landesamt für Datenschutzaufsicht. EU-Datenschutz-Grundverordnung (DS-GVO) - Das BayLDA auf dem Weg zur Umsetzung der Verordnung - Umgang mit Datenpannen - Art. 33 und 34 DS-GVO. `https://www.lda.bayern.de/media/baylda_ds-gvo_8_data_breach_notification.pdf`, 2016. [Online; accessed March 2018].

[2] Catalin Cimpanu. US Congress Passes CLOUD Act Hidden in Budget Spending Bill. `https://www.bleepingcomputer.com/news/government/us-congress-passes-cloud-act-hidden-in-budget-spending-bill/`, 2018. [Online; accessed March 2018].

[3] J. DeBlasio, S. Savage, G. M. Voelker, and A. C. Snoeren. Tripwire: inferring internet site compromise. In *Proceedings of the 2017 Internet Measurement Conference*, pages 341–354. ACM, 2017.

[4] Derric Gilling. Capturing AJAX API Requests From Arbitrary Sites With a Chrome Extension. `https://dzone.com/articles/how-we-captured-ajax-api-requests-from-arbitrary-w`, 2017. [Online; accessed March 2018].

[5] Dr. Thomas Schwenke. MailChimp, Newsletter und Datenschutz. `https://drschwenke.de/mailchimp-newsletter-datenschutz-muster-checkliste/`, 2016. [Online; accessed March 2018].

[6] Dr. Thomas Schwenke. Datenschutz und ePrivacy 2018 – Änderungen für Onlinemarketing, Tracking und Cookies. `https://drschwenke.de/datenschutz-eprivacy-online-marketing-cookies/`, 2017. [Online; accessed March 2018].

[7] EUROPEAN PARLIAMENT. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). `http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=en`, 2016. [Online; accessed March 2018].

[8] J. Farrell and P. Klemperer. Coordination and lock-in: Competition with switching costs and network effects. *Handbook of industrial organization*, 3:1967–2072, 2007.

[9] Google. IP Anonymization in Analytics. `https://support.google.com/analytics/answer/2763052`. [Online; accessed March 2018].

[10] J. R. Mayer and J. C. Mitchell. Third-party web tracking: Policy and technology. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 413–427. IEEE, 2012.

[11] Netzpolitik.org. Cambridge Analytica: Was wir über „das größte Datenleck in der Geschichte von Facebook" wissen. `https://netzpolitik.org/2018/cambridge-analytica-was-wir-ueber-das-groesste-datenleck-in-der-geschichte-von-facebook-wissen/`, 2018. [Online; accessed March 2018].

[12] Slaughter and May Legal Services. New rules, wider reach: the extra- territorial scope of the GDPR.

https://www.slaughterandmay.com/media/2535540/new-rules-wider-reach-the-extraterritorial-scope-of-the-gdpr.pdf, 2016. [Online; accessed March 2018].

[13] M. Wachs, Q. Scheitle, and G. Carle. Push away your privacy: Precise user tracking based on tls client certificate authentication. In *Network Traffic Measurement and Analysis Conference (TMA), 2017*, pages 1–9. IEEE, 2017.