GDPiRated – Stealing Personal Information On- and Offline

Matteo Cagnazzo[1], Thorsten Holz[2] and Norbert Pohlmann[1]

[1]Institute for Internet-Security, University of Applied Sciences Gelsenkirchen, Gelsenkirchen, Germany {lastname}@internet-sicherheit.de
[2]Horst Görtz Institute (HGI), Ruhr-Universität Bochum, Germany
thorsten.holz@rub.de

Abstract. The European *General Data Protection Regulation* (GDPR) went into effect in May 2018. As part of this regulation, the *right to access* was extended, it grants a user the right to request access to all personal data collected by a company about this user. In this paper, we present the results of an empirical study on data exfiltration attacks that are enabled by abusing these so called *subject access requests*. More specifically, our *GDPiRate attack* is performed by sending subject access requests (as demanded by the GDPR) with spoofed recipient addresses either in the on- or offline realm. Our experimental results show that entities accepting and processing offline requests (e.g., letters) perform worse in terms of ensuring that the requesting entity is the correct data subject. The worrying finding is that affected organizations send personal data to unverified requests and therefore leak personal user data. Our research demonstrates a novel attack on privacy by abusing a right the GDPR tries to protect.

Keywords: GDPR · Privacy · Offensive Security.

1 Introduction

On May 25, 2018, the General Data Protection Regulation (GDPR) went into effect in the European Union. Its major goal is to harmonize privacy protection mechanisms across the European Union (EU) and enable users to exercise their rights, whenever and wherever data of them is processed. On- and offline services provided to European citizens are affected by these changes and required to adopt this regulation. An important aspect of the GDPR is the *right to access*, which grants a user the right to request access to all personal data collected by a company about this user via a so called *subject access request* (SAR).

While these changes are generally considered positive in terms of privacy and transparency for users, little research has been done on how these new mechanisms could be exploited by an adversary trying to gather personal information. In this paper, we describe an empirical case study we conducted in the year 2018 using rather simple techniques to exfiltrate personal data out of municipal, healthcare, and other providers that process sensitive data. Broadly speaking,

we send SARs to a company and request access to data belonging to a victim user, an attack we call *GDPiRate attack*. Due to the private nature of the data we exfiltrated and the fact that most of the services we examined are identity- and location-based in terms of who uses the platforms, we decided to conduct only a small case study with data from the authors to not harm any person. We also took special measures to limit the potential impact of our analysis and to perform the attack in a responsible way. The exfiltrated data could especially be misused by attackers who are trying to *dox* other users or want to conduct targeted (spear)phishing attacks in an advanced persistent threat (APT) or a fraud scheme. Our experiments hence required special handling of sensitive data.

Our experimental results show that despite all the positive effects that the GDPR had on privacy in Europe, a new attack surface is made available due to SARs and the fines that organizations face if they do not respond within one month. We show that a lot of organizations do not seem to have proper protection and authentication mechanisms in place to verify if a subject access request indeed belongs to a legitimate and authorized user. More specifically, 10 out of 14 spoofed SARs were successful and we were able to obtain access to sensitive information. Whenever services or organizations provide a mix of online and offline data and do not have a dedicated process dealing with SARs, it appears that these services are susceptible to our GDPiRate attack. The short timeframe which is given to companies to react to SARs is a positive effect to transparency of users, but also facilitates our attack. It seems that only a small number of companies have proper authentication processes for SARs in place to determine if a legitimate user or an attacker contacts them via mail, fax, or email.

In summary, our paper makes the following contributions:

- In a case study, we introduce a new attack on how the GDPR might help identity thefts, doxers, and fraudsters to gain access to personal information. Our empirical results indicate that spoofed subject access requests are often successful in practice (10 out of 14 cases) given that companies to not perform enough checks to verify the identify of the source of a SAR.
- We analyze the collected data and categorize the impact using the Common Vulnerability Scoring System (CVSS) based on the accessible data.
- We discuss both operational and strategic mitigations for the vulnerabilities we found. We hope that our research raises awareness of this problem and that more companies start to implement proper mechanisms to verify the identity and source of SARs.

2 Related Work

According to a study by the Ponemon Institute [31], 52 % of companies said they are GDPR-ready before the deadline in May 2018. The same study found that 50% of companies in the health sector and 47.2% in the public sector expected to

be GDPR compliant before this date. Contrary to this study, we found that major flaws exist in the operational section of privacy implementations, especially in terms of subject access request handling.

There have been many studies on privacy policies, presumably because they are the primary factor in terms of transparency [2, 21, 23, 28, 40]. There is much work on how privacy policies are perceived by users, what they disclose, and how accessible the information is to users. For example, Degeling et al. examined popular websites in the European union and their implementation of the GDPR [8]. Linden et al. reviewed privacy policies pre- and post-GDPR and found overall positive improvements to privacy rights of citizens in the EU [30].

Another important factor are cookies and web tracking in general [11, 12, 44]. For example, Urban et al. showed that cookie providers generally comply with GDPR, but not within an appropriate timeframe [52].

Our GDPiRate attack is a general attack not exploiting a technical flaw, but rather an organizational and human protocol flaw. A technical analogy would be to visit a website and by changing the user agent string of the web browser, an attacker would gain access to personal data on this website. A similar approach has been taken by Gruss et al. [22], where the technical use-after-free vulnerability of memory unsafe languages like C/C++ was generalized to "any environment and situation where resources can be silently exchanged". While Gruss et al. wait for resources to be freed or released, GDPiRate uses the information while it is still in use of the legitimate user.

Concurrently and independently, Di Martino et al. [32] performed a similar study in Belgium and showed that the attack is not only feasible in Germany but also in Belgium. This highlights the fact that our attack seems to be a bigger European problem which needs to be solved, so the privacy of individuals is better protected.

3 Background

In this section, we provide a brief overview and introduction to the GDPR and its history. Furthermore, we cover attack techniques such as social engineering, phishing, and doxing.

3.1 General Data Protection Regulation (GDPR)

The need to protect the privacy of individuals and still be able to share data across the EU has been an European effort for more than twenty years. In 1995, the Data Protection Directive 95/46/EC was created to regulate the processing of personal data and harmonize European data processing. Directives are not directly applicable to members of the European Union. Each state has to adopt an individual implementing act. Therefore, the harmonization of European privacy laws failed. Recital 9 of the GDPR explicitly points out that implementations differed widely, which resulted in a complex privacy law landscape across Europe.

In 2016, the European Parliament and Council came to an agreement on a new data protection law following a four year proposal and trilogue period. The Council and Parliament decided that the General Data Protection Regulation (GDPR) will be fully enforceable throughout the EU after a two year post-adoption grace period. The aim of the GDPR is to protect citizens of the EU from privacy violations and data breaches. The GDPR's integral goal is to harmonize privacy laws all over Europe. In addition to the GDPR, the ePrivacy regulation is currently passing the EU's legislative process to complement the GDPR.

One of the key changes of the GDPR in accordance to harmonizing the law landscape on privacy is an increased territorial scope (Art. 3 GDPR). It applies to all entities processing personal data of subjects residing within the EU. It is not important where the processing entity is located. Furthermore penalties can be fines up to 4% of annual global turnover or 20 million euro. This fine is applicable if user consent is not sufficient or other core privacy principles are violated. User consent is another key change of the GDPR because the conditions became stronger. Data processors must provide the request for consent in an easily accessible form with purpose of processing directly attached to this form. Withdrawal of consent must be as easy as it is to give consent (Art. 7 GDPR).

Data must be protected by design and default. Article 25 of the GDPR states that entities need to *implement appropriate technical and organizational measures [...] designed to implement data protection principles[...] in an effective manner [...] taking into account the state of art* [16]. This means that processing entities must ensure that private data is not publicly available without the user's consent. Higher protection standards must be implemented for sensitive categories of data like religious or health data (Art. 9 GDPR) [17].

Another integral change for data processors is the *right of access*. Every user has the right to access their personal data which is collected and processed, according to Article 15 [15]. If they are requesting this in an electronic form, the answer must be provided in an electronic manner as well. First access to their personal data is to be provided free of charge by the person responsible with a copy of the data subjects available data. If the request is made via postal mail, the response should also be sent via mail. Recital 74 states that the identity of the applicant must be verified so that only the person whose data is processed has access to the data [15]. According to Article 12 [18], a processor can request additional information to confirm the identity of the data subject requesting the information. The GDPR states companies "should use all reasonable measures to verify the identity of a data subject who requests access", to make sure they do not disclose data to the wrong person. This identification process is supposed to add a layer of security. If a data processor is not complying to such a request in a reasonable timespan, he can be fined according to the fines mentioned above.

According to Art. 9 of the GDPR, the following types of data are in a special category and processing of this data is always prohibited unless explicit consent is given by the data subject [17]:

- racial or ethnic origin
- political opinions

- religious beliefs or other beliefs of a similar nature
- trade union membership
- physical or mental health or condition
- sex life and sexual orientation
- genetic data and biometric data

We use healthcare data to describe special categories in detail, the other categories follow a similar scheme. According to Art. 4 of the GDPR, healthcare data is defined as *personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status;* [14]. This data is sensitive and needs specific protection from unlawful processing and data leaks. A report by the EC confirms that respondents from the EU are concerned about applications tracking their activities and vital data [13].

Even though the data is sensitive and needs more protection, a study by the compliance analytics company Protenus [42] found that in 2017 there were 477 data breaches reported to the United States Department of Health and Human Services, which is a slight increase from previous monitoring periods. The main reasons for these breaches are hackings and insider threats [42]. According to Fuentes [19], attackers found multiple ways to profit from stolen medical health records for example by misusing data to get medical prescriptions or perform identity theft.

Apart from official medical institutions being targeted, there is a large number of healthcare online platforms and applications which closely monitor users' vital parameters. The applications range from weight tracking to suicide prevention, from online platforms where types of donors are managed. They all have in common that they hold and process sensitive data which is related to (mental) health of users. Papageorgiou et al. [38] found that mobile health applications do not follow well-known best practices and guidelines, or comply to legal restrictions. Rasthofer et al. [43] analyzed multiple mutual agreement tracking applications and found that the state of security in these applications is worrisome.

Alongside badly coded applications and adversaries, some companies that track data share anonymized datasets. For example, fitness trackers that track heartrate and geolocation offer publicly available data, which can be deanonymized because too much private information is embedded in the anonymized data [24]. This data is very specific for individual users, causing more pretext information for (spear)phishing attacks [5].

Starting our research of GDPiRate with healthcare services, we realized that there are more entities which are vulnerable to our attack. Especially credit bureaus, address dealers, tenant information services, and transport services could be vulnerable. Hence we decided to include such companies into our study, even though they are not connected to the medical landscape. Nevertheless, they also handle sensitive and private data which is categorized as data with increased protection requirements.

3.2 Social Engineering

In the information security context, social engineering is an attack technique with the goal of influencing potential victims to disclose information or perform a specific action. Attackers do not exclusively use technical means; social manipulation is at the forefront of social engineering. Criminals often make use of social, established behavior. Holding open an access-restricted door for another person would be such an exploitable behaviour. The information or actions received are often just a means to an end, and in most cases are used to enter an information processing system or to gain access to an asset within a company. The manipulation can affect trustworthiness, integrity, or availability. Deception and manipulation are the basics of social engineering-based attacks. Against private individuals, social engineering campaigns are usually conducted to impersonate identity theft. For example, the user is directed to fake web portals to enter credentials, such as username, password, or banking details. Individuals also become victims of social engineering to steal access to company information.

Current technical countermeasures are usually ineffective against these attacks. Companies have little information about what data employees share on the Internet. In addition, many victims of social engineering attacks feel that they are good at detecting these attacks, but in practice they are not. An increased risk comes from *Unintentional Insider Threats* (UIT). The CERT Insider Threat Team defines a UIT as follows: "An Unintentional Insider Threat is (1) a current or former employee, contractor or business partner (2) who has or has authorized access to an organization's network, system or data Action or omission without malicious intent (4) will cause harm or substantially increase the likelihood of future serious damage to the confidentiality, integrity or availability of the organization's information or information system." These insider threats, whether intended or not, are considered to be one of the greatest risks to operational and organizational security. The two threats are different in motivation, indication, and other differences. Therefore, it is important to study and understand these threats in detail.

A common route of infection is phishing which is explained in the following section. Another emerging trend in social engineering is the exploitation of machine-learning-based insights to target users individually with phishing links.

3.3 Phishing

Phishing is a neologism of "fishing" and an act of deception. For the first time, the term was used in 1996 in the context of a theft of America Online (AOL) accounts [36]. Previously, swindlers used classic scams to help people to save their data or belongings. False messages promising money go back to the nineteenth century. Suspected Spanish prisoners wrote letters to wealthy US citizens. The author is always in prison as a political prisoner, but has a large amount of money to dispose of as soon as he is released. To obtain the dismissal, the attacker only needs some money. It is unclear how much damage has actually been done, because even then the prosecution was a problem [50].

In phishing, an attacker uses a fake e-mail, short message, or website to trick a victim into revealing non-public information. Often, passwords or bank details are stolen, but personal information can also be lost. One of the most common problems with phishing attacks is one that is prevalent in any deception: the victim often does not realize that he has been the victim of fraud. Only the abuse of information often leads to the victim noticing that they were deceived. For example, an American study shows that more than 1 million children in the US have already been the victims of identity theft. The resulting economic damage amounts to $ 2.6 billion in damage, which is transferred directly to the families [39].

Phishing scales very well because there are hardly any costs for sending an e-mail. An attacker needs an e-mail account and an Internet connection, then they can start attacking. In its specification, the SMTP protocol offers no possibility to authenticate a sender [29]. Although e-mail is the primary means of phishing, phishing is also available by phone (*vishing*), SMS (*smishing*), and social networks (*Snishing*) [54]. Especially in social networks, it is more complex for an attacker to scale, because there is trust among users. Attackers often resort to creating false customer support accounts and then phishing links in the comment columns. This procedure is called "angler phishing": a study by Proofpoint shows that about 30% of all known market accounts are a form of fraud and security risk [41]. These fraudulent acts are different. From the theft of account data to click fraud to the distribution of pornographic material. This means that customers of the affected brands lose data, while it also leads to a reputation loss for the brand.

The phishing link is distributed by an attacker in an arbitrary way, for example in social networks or via an e-mail. It should be noted that there is no prevailing modus operandi for phishing, but an adversary can choose from a variety of possible procedures. An attack, however, usually consists of three steps: 1) setting up the attack, 2) carrying out the attack, and 3) monetizing the information. In the first phase of the attack, the phishing emails are created and the page on which the users submit their information is set up. In addition, the circle of victims must be defined, for example, by collecting e-mail addresses. In this phase, spear phishing differentiates itself from phishing [3]. Spear phishing often contains *pretexting*, which represents the "phisher" as a trusted third party and tailored to individual preferences of individual victims. Information from the private or professional environment of the victim is the key to success for this type of manipulation. Using this information enormously increases the probability of success of an attack, even when users are trained and have increased awareness [6]. Advanced Persistent Threats (APTs) can include spear phishing as an initial compromise vector, such as Operation Pawn Storm or TG-4127 [46] [51].

3.4 Doxing

Doxing is an attack where private information of a victim is released online [45, 48]. This attack is technically unsophisticated, but it can cause serious harm to its victims not just online but also offline, as quantitative [49] and qualitative

research [10] shows. Hacking technologies or services, on the other hand, involves a certain skillset of an attacker to cause social harm to a victim. Doxing is, technically speaking, an unsophisticated release of documents. Prior to releasing information, the adversary must somehow access sensitive data about the victim.

An interesting case study is for example the release of e-mails of the Democratic Party during the 2016 US elections. The Mueller report [35] states that "a Russian intelligence service conducted computer -intrusion operations against entities, employees, and volunteers working on the Clinton Campaign and then released stolen documents". A sophisticated hacking effort took place before the emails and other materials were released by the two personas "Guccifer2.0" and "DCLeaks" (and later on Wikileaks).

Especially in Germany, doxing gained attention and awareness during 2019 due to mass doxing of German politicians [4], where a twenty year old suspect was accused and arrested for collecting and releasing partly public information of German politicians and celebrities. This leak was also partly politically motivated even though no state actor was involved. The suspect said he was "angry over politicians" [27].

Note that doxing is not always hostile. As Chen et al. [7] find, there are two main categories of doxing: hostile and social doxing. About half of the people studied by Chen et al. said they perform social doxing to fulfil social needs, for example obtaining social data or relationship statuses. In contrast, the other half of the respondents of this study used doxing in a hostile way.

4 Approach

Our study on potential ways to abuse subject access requests builds on simple techniques, no cryptography or technical concept was broken to achieve private data exfiltration from organizations and municipal administrative authorities. In the following, we describe our approach in more detail.

4.1 Ethical Considerations and Responsible Disclosure

Given the sensitive nature of this study, we first discuss both ethical considerations and the way we handled responsible disclosure. Neither Ruhr-Universität Bochum nor the Westphalian University for Applied Sciences have an Institutional Review Board (IRB) for computer science. Our work was conducted according to ethical best practices [1, 9, 33] and privacy laws. We used synthetic information wherever possible and handled the required data securely in terms of access control and storage. If data was sent to us via normal postal mail, the letters were stored in a locked container inside a locked office. Only researchers participating in this paper had access to this container and if non-synthetic data was used, it was personal data belonging to one of the authors. If data was sent to us electronically, it was sent to private or synthetic mailboxes.

We stored electronically received data on encrypted systems only. All private data was sent to spoofed mailboxes. Only the owner of the requested data had

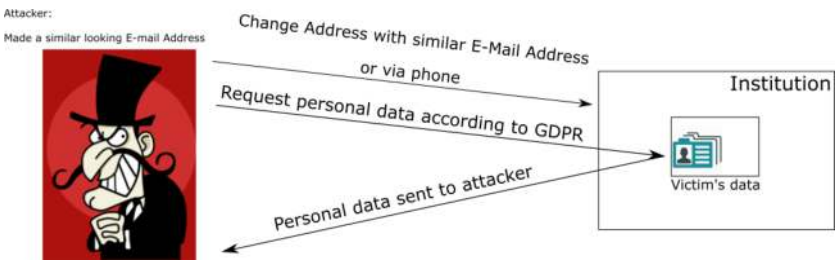


Fig. 1. High-level attack overview for *GDPiRate*

access to these mailboxes. Furthermore, we did not report any of the vulnerable organizations to a third party (e.g., a data protection authority). We anonymized the identity of the vulnerable organizations in this paper to minimize fraudulent activities based on our research. In addition, we modified individual proofs of identities before sending the SAR to an organization, therefore it should be mentioned that no official government documents were altered for this research. By the time of publication, we have contacted the affected organizations and informed them about the outcomes of our research. We gave each affected organization individual advice on how to improve their handling of SARs. In this paper, no information about the identity of affected organizations is included to minimize potential harm and damage related to the reputation of the affected organizations.

4.2 Attacker Model and Approach

We use an active adversary to interact with institutions and companies such that we can study how these entities react to spoofed SARs. The general goal of our attack is to gather personal information that is potentially sensitive.

A high-level overview of the attack flow is shown in Figure 1, we call it the *GDPiRate attack*. The attacker needs to know with which e-mail address or real-world address a victim is registered at a certain service or organization. For example, she can figure this out with open-source intelligence. If she knows the e-mail address, she registers a similar-looking address (e.g., if the victim's e-mail address is `example@example.com`, the attacker would register `exmple@example.com` or another perturbation of the e-mail address). She then contacts the institutions support via e-mail to change her residing address. If they reject this change, she will call them with a spoofed number either of the victim or the city the victim is residing in, depending on the level of open-source intelligence she has acquired. After she initiated the address change, she waits a few days until she requests "her" data according to Art. 15 of the GDPR. In return, she will either receive the data directly to her spoofed e-mail address or via postal services to her spoofed street address. If she only knows the name of the victim, she can try to request data from the services or organizations just using a spoofed address. If she knows the street address of her victim, she can use offline channels

and a similar strategy over which a SAR can be made. All an attacker needs to know is the last place of residence. She can say she moved and requests her data now to the new address. This will lead to data sent out to the attacker and the integrity of the data being changed. For both types of this attack, she needs to put up a sign on her mailbox.

To evaluate how effective the GDPiRate attacks is in practice, we performed an empirical study in which the authors registered an account with synthetic data. In the evaluation phase, we spoofed our own identity as an attacker to request this data over another channel. The text we used is available in the Appendix of this paper, one version in German and one in English. As part of this experiment, we also requested data with correct identity information.

5 Results

We contacted 14 services and evaluated how they react to our spoofed SARs, basically to test whether they implement proper mechanisms that only entitled and correctly identified users can gain access to private information. We focused on Germany since municipal data can only be requested for a real person and not synthetic data. We evaluated whether these 14 parties have a privacy policy and send out personal data to a spoofed recipient. To perform our SARs, we mainly sent out e-mails or used a fax machine as described in Section 4.2. Table 1 provides a detailed overview of our findings.

We verified our results in a control phase by requesting data about the same "victim" to the correct address and analyzed whether different data was sent. This was done to prevent an assumption that the requested entity will look up the victim at the spoofed address and if he is not registered at this address, they will just send out that a response that states that no data is stored.

Our study shows that all the requests we made were answered during the required time period. In our control phase, the requests were answered in the appropriate time frame as well. One vendor needed further clarification on which data we wanted to access in both time periods. All services we tested had a privacy policy in place and it was possible to interact with them via e-mail or telephone. Of our fourteen tested services, only two services reacted to the spoofed address by sending us a letter back requesting verification of our spoofed address. The other two services that are stated as "no" in Table 1 stated they do not process personal data.

All other companies sent out personal information, without asking for additional verification of the spoofed address and identity. During the control phase, we were able to verify if companies had other "identification processes" in place, for example verifying the identity via postal address. No company we tested sent out other data to the real address. We believe no verification of the identity is in place, therefore all these services violate the GDPR since they send out personal information to spoofed recipients and are therefore vulnerable to the GDPiRate attack. An overview of our results can be seen in Figure 2.

Service	Purpose	Location	Has Privacy Policy	Sent Out Personal Data to spoofed recipient	Particularly sensitive data	Responded in time
Medical Institution 1	Donation service	Germany	Yes	Yes	Yes	Yes
Medical Institution 2	Donation service	Germany	Yes	Yes	Yes	Yes
Broadcasting Service 1	Broadcasting media	Germany	Yes	Yes	Yes	Yes
Postal Service 1	Postal Services	Germany	Yes	Yes	No	Yes
Travel Service 1	Transportation Services	Germany	Yes	Yes	Yes	Yes
Financial Service 1	Credit Bureaue	Germany	Yes	Yes	Yes	Yes
Financial Service 2	Creditbureau	Germany	Yes	Yes	Yes	Yes
Financial Service 3	Credit Bureau	Germany	Yes	Yes	No	Yes
Financial Service 4	Credit Bureaeu	Germany	Yes	Yes	Yes	Yes
Financial Service 5	Credit Bureau	Germany	Yes	No	No	Yes
Financial Service 6	Credit Bureau	Germany	Yes	No	No	Yes
Financial Service 7	Credit Bureau	Germany	Yes	No	No	Yes
Adress Service 1	Adress Broker	Germany	Yes	Yes	No	Yes
Government Service	Federal Bureau	Germany	Yes	No	No	Yes

Table 1. Overview of results

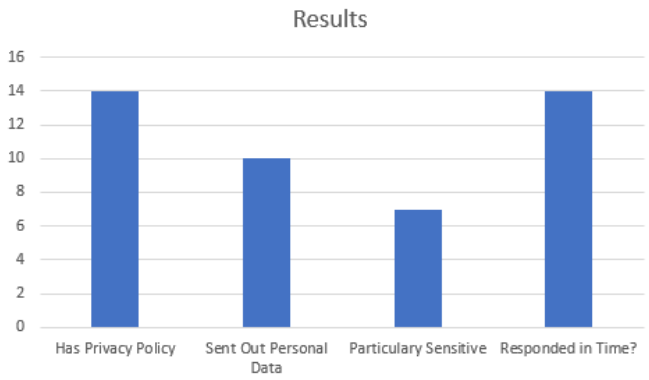


Fig. 2. Results of empirical study of GDPiRate attack.

We found that an attacker can possibly steal private and sensitive information with a success of approximately 71% (10 out of 14 cases). In half of the cases, an attacker obtains sensitive data like financial or healthcare information. For example, it is possible to obtain access to health data like blood type, HIV status, or if blood has been donated and where. This is critical private data and should never be handed out to a spoofed request. Our study shows that it is easily possible to steal data from various services to do so with a spoofed mail or postal address according to our attack scheme shown in Figure 1. Here is an exemplary list of data we were able to obtain:

- Health data
- Complete bank details (IBAN/BIC)
- Credit scores
- Business registration history
- Debt collection information
- Complete train travel history of the last 12 months
- Date of birth

In two cases, we were able to observe that the integrity of the data changed. Our request to the spoofed address changed or added the spoofed address in the control SAR response.

Especially health and financial data is critical since they clarify as special data category according to the GDPR. We also obtained other types of data that can be used for fraudulent activities by criminals and should therefore be mitigated. Every bit of analog information is valuable for an attacker because it can be referenced forever. For example, if a person has a heart condition this is unlikely to change and if a person lived at a certain address paying with a certain bank account, this data will likely not change either. Digitally generated data can be used and deleted later, but every data point that has its source in the real world cannot be deleted electronically.

We were able to access the whole train travel history including payment details, real address and more metadata like mobile numbers, employment information, and profession of one of the authors. The travel data we were able to send to a spoofed address is depicted in Figure 3. Some data is redacted due to privacy reasons. This data was sent to the spoofed address. This is especially critical in doxing scenarios and a starting point for every information-based attack such as (spear)phishing.

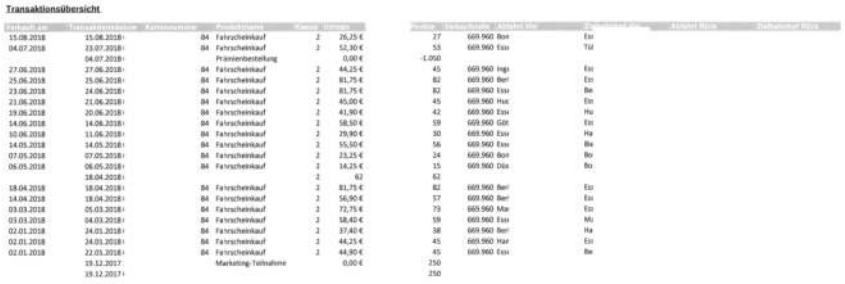
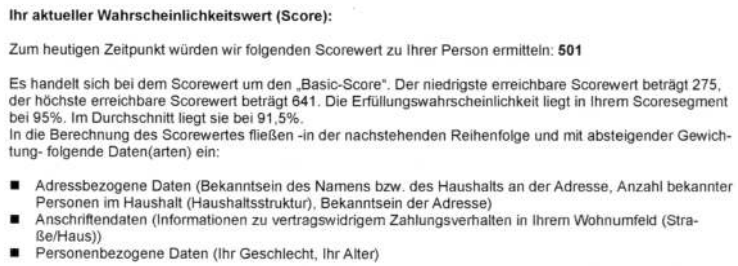


Fig. 3. Travel history sent to spoofed mailbox

We used the CVSS framework to rate our vulnerability. Of course, CVSS is normally used to describe software vulnerabilities, but we describe our attack using this framework as well [34]. We get a score of 8.2 (High) for GDPiRate. The CVSS string is:

`CVSS:3.0/AV:N/AC:L/PR:N/UI:R/S:C/C:H/I:L/A:N/E:F/RL:W/RC:C/CR:H/IR:H/AR:M`.

We defined the attack vector as *network* because we can trigger the attack via mail, letter, or fax from anywhere in the world. This is not OSI layer 3 as defined by CVSS, but it is the option that is closest to our attack. All we need is an endpoint to send our SAR to and it is possible to trigger the attack from anywhere in the world as long as the spoofed mailbox is located in the country of the victim. In terms of attack complexity, we rank the vulnerability with *low* because an attacker just needs a name and can send spoofed SARs to the entities. With a success probability of 71% in our study, it is likely that the attack is successful and an adversary can collect personal information in a spoofed mailbox either on- or offline. The attacker needs no privileges to trigger the attack. She can just start the attack and does not need any specific privileges. If an entity is unsure about the identity of the attacker, they will likely request additional verification. A successful attack is dependent on the operator processing the SAR. Therefore we assume that we need user interaction for our attack to be successful in the wild. Our attack strategy does not verify whether a



Ihr aktueller Wahrscheinlichkeitswert (Score):

Zum heutigen Zeitpunkt würden wir folgenden Scorewert zu Ihrer Person ermitteln: **501**

Es handelt sich bei dem Scorewert um den „Basic-Score". Der niedrigste erreichbare Scorewert beträgt 275, der höchste erreichbare Scorewert beträgt **641**. Die Erfüllungswahrscheinlichkeit liegt in Ihrem Scoresegment bei 95%. Im Durchschnitt liegt sie bei 91,5%.
In die Berechnung des Scorewertes fließen -in der nachstehenden Reihenfolge und mit absteigender Gewichtung- folgende Daten(arten) ein:

- Adressbezogene Daten (Bekanntsein des Namens bzw. des Haushalts an der Adresse, Anzahl bekannter Personen im Haushalt (Haushaltsstruktur), Bekanntsein der Adresse)
- Anschriftendaten (Informationen zu vertragswidrigem Zahlungsverhalten in Ihrem Wohnumfeld (Straße/Haus))
- Personenbezogene Daten (Ihr Geschlecht, Ihr Alter)

Fig. 4. Credit score sent to spoofed mailbox

Personendaten:

Unser Unternehmen hat zu Ihrer Person folgende Daten gespeichert:
Vor- und Nachname:
Straße und Hausnummer: Changed Spoofed Address
PLZ und Ort:
Geburtsdatum:

Fig. 5. Integrity of data is altered

real user interacts with the SAR or everything is automated. The scope changes because we are requesting victim information and data over a third party which processes the spoofed SAR. If we want to request information on victim A, we ask organization B impersonating victim A.

The GDPiRate attack breaks confidentiality of the user data which is especially critical if health, financial, or other personal data is obtained, which is the case in every second spoofed SAR response we obtained. Figure 4 shows a credit score we were able to obtain to a spoofed mailbox.

The integrity of the data can be altered if we can change addresses beforehand, which was the case in two cases. Figure 5 shows a response we received where the spoofed address was used instead of the real author's address. The figure is redacted for privacy reasons. We could not monitor a change of the availability of data.

The data we gathered during our case study is very valuable to an attacker that wants to phish, spear phish, social engineer, or dox a victim. An attacker could use the data she obtains through the GDPiRate attack to attack a user in a private or business context. The attack is very targeted but can have devastating effects on single users and organizations. Note that our attack gives away very specific information about single targets, while being totally unsophisticated. All an attacker needs to do is set up a spoofed mailbox electronically or in the real world. Reconnaissance and information gathering is an important step in successful social engineering attacks. Especially individual, non-generic information like the information we gather with GDPiRrate can provide valuable information to an attacker because she has correct information about her victims.

6 Mitigation

Mitigating our attack is not trivial given that the underlying problem is not on the technical level. We combine off- and online identities with on- and offline communication channels and it turns out that these cannot be easily verified in the same system apparently. In most cases, a holistic approach to the connection and verification of on- and offline identities is needed, but there is unfortunately no out-of-the-box solution for this at the moment. Another facet is that the detection of spoofing is a hard problem even in the digital realm. A recent study [26] finds that for example most mail providers offer spoofing detection, but still deliver some spoofed mails to inboxes. If an email is from an unverified sender, only nine out of 35 mail providers provide visual clues for the end user.

The main problem is that real-world information can be referenced forever and is therefore valuable for all kinds of attackers and attacks. This is not a new insight, but with the informative duty that organizations now have under GDPR, it is important that users and organizations realize that everything that is not a *cryptographic secret* cannot be used as an authorization factor because one cannot expect it to be secret. If cryptographic secrets are not used, but rather something public such as a name and address is used, this can be abused by an adversary as the entry point to a variety of personal information as shown with GDPiRate. As we discussed in the previous section, the attack is easy to exploit and since spoofed emails are a common threat, we cannot rule out that this attack has not been exploited so far in the wild.

Currently there are several projects that try to map an offline identity to an online user or application, e.g., in ecommerce [20] or municipal data [25]. If these research projects are successful, this could be one way to map municipal data to an off- and online persona at least. This could be applied by private data brokers and SAR respondents as well to be sure that an offline entity requesting information can be verified as the digital identity to which the data belongs.

Companies should review their implementation and processes of how user data is accessed and forwarded under a SAR. These processes seem to be implemented incorrectly in many cases. To implement these processes correctly requires organizations to look past the border of securing data online, but also verify the security and privacy of processes and non-digital communication channels in organizations. In the short term, companies should implement precautions and extra steps to avoid sending out private data to any attacker.

The attack can be mitigated if SARs are made within a dedicated web portal or application function. The user identifies via a username/password/2FA combination and is authenticated. The user requests his data using a dedicated function in the application. This prevents the GDPiRate attack most efficiently, because there is a fixed process to gain access to personal data according to the GDPR. If organizations do not have a web portal or application they need to make sure that they provide a robust mechanism to verify the identity of the requester.

One way to mitigate the attack could be to ask for additional information if a data point that is used to identify a user is wrong, for example the recipient

address does not conform with the data an organization has. During our case study, one of the companies performed such as check when we requested data with a fake address and persona. In particular, this company obtains data directly from banks where it is required to open a bank account with an identity card. They have a verified data point to which they can fall back if a spoofed SAR is sent to them and request additional information before they send out private data. Such processes could be implemented in other companies as well.

GDPiRate can generally be mitigated by making staff more aware of how privacy can be broken in the analogous world. If staff sees a wrong data point submitted in a SAR, they should request additional data to identify the requesting entity. For example voice recognition on the phone or a holistic identity and verification framework could be used. If voice recognition techniques are used, it should be mentioned that there are ways to break them too as an authentication mechanism, but it would significantly increase the attack complexity for a successful attack [47].

7 Conclusion and Future Work

Mitigating and avoiding GDPiRate is essential for organizations that process personal data, otherwise they could send out private data to a false requestor or get fined according to GDPR guidelines because they did not implement countermeasures to false requests. Even with the GDPR in place and demanding quick response times to SARs, just sending out personal information of a user to an attacker who only uses a spoofed address and another communication channel is clearly problematic. Especially in times where politically-motivated doxing and targeted attacks against organizations and governments using (spear)phishing attacks are on the rise.

Our work has limitations due to the relatively small sample size of fourteen data processors where we requested data with a SAR. We had to make sure that GDPiRate works, while providing integrity of the data we used, since in most cases we were not able to use synthetic data beforehand and used real-world data of one of the authors. Doxing as a service and the commercialization of doxing made it easier to conduct harassment online. An attacker can use a dox-for-hire service to conduct research and compile information about potential victims from as low as $50 [53]. A dox is more valuable the more private, previously undisclosed data it contains. Therefore organizations need to make sure that they do not provide anyone else than the real owner personal data access to the data (as also stated by the GDPR).

Further research should focus on preventing these attacks and to bridge the gap between online and offline identity, while preserving as much privacy as possible for each individual. This work is not intended to pave the road for laws like a planned Austrian law ("Bundesgesetz über Sorgfalt und Verantwortung im Netz - SVN-G"), where users are required to register before they can connect and interact with the Internet and therefore abandoning online privacy completely [37].

In conclusion, it becomes more apparent that a vast research field emerges at the intersection of digital and analog processes and the transformation between an analog process to a digital process is complex in terms of security and privacy. Private data that is stored in an organization needs to be stay private and must be accessible only to the authorized data owner. If the organization offers a digital authentication, it should use this mechanism to provide data only to authenticated users. This would prevent an unauthenticated, spoofed attack like GDPiRate is.

Acknowledgements This work is partly funded by the Federal Ministry of Education and Research in Germany (Grant.Nr: 16SV7775 and Grant.Nr: 16KIS1016).

References

1. Bailey, M., Dittrich, D., Kenneally, E., Maughan, D.: The Menlo Report. IEEE Security & Privacy **10**(2), 71–75 (Mar 2012)
2. Bélanger, F., Crossler, R.E.: Privacy in the digital age: A review of information privacy research in information systems. MIS Q. **35**(4), 1017–1042 (Dec 2011)
3. Benenson, Z., Gassmann, F., Landwirth, R.: Unpacking spear phishing susceptibility. In: International Conference on Financial Cryptography and Data Security. pp. 610–627. Springer (2017)
4. Bundeskriminalamt: Festnahme eines Tatverdächtigen im Ermittlungsverfahren wegen des Verdachts des Ausspähens und der unberechtigten Veröffentlichung personenbezogener Daten. Online (Jan 2019), `https://www.bka.de/DE/Presse/Listenseite_Pressemitteilungen/2019/Presse2019/190108_FestnahmeDatenausspaehung.html`
5. Cagnazzo, M., Pohlmann, N.: Using geolocation data as a threat enlargener for social engineering attacks. In: DACH Security (2019)
6. Caputo, D.D., Pfleeger, Lawrence, S., Freeman, J.D., Johnson, M.E.: Going spear phishing: Exploring embedded training and awareness. IEEE Security & Privacy **12**(1), 28–38 (2013)
7. Chen, M., Cheung, Yue, A.S., Chan, K.L.: Doxing: what adolescents look for and their intentions. International journal of environmental research and public health **16**(2), 218 (2019)
8. Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F., Holz, T.: We Value Your Privacy... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In: Network and Distributed Systems Security (NDSS) (2018)
9. Dittrich, D., Kenneally, E.: The menlo report: Ethical principles guiding information and communication technology research. Tech. rep., US Department of Homeland Security (2012)
10. Douglas, D.M.: Doxing: a conceptual analysis. Ethics and information technology **18**(3), 199–210 (2016)
11. Englehardt, S., Narayanan, A.: Online tracking: A 1-million-site measurement and analysis. In: ACM SIGSAC Conference on Computer and Communications Security. pp. 1388–1401. CCS'16 (2016)
12. Englehardt, S., Reisman, D., Eubank, C., Zimmerman, P., Mayer, J., Narayanan, A., Felten, E.W.: Cookies That Give You Away: The Surveillance Implications of Web Tracking. In: International Conference on World Wide Web. pp. 289–299. WWW'15 (2015)

13. European Commission: Special eurobarometer 431: Data protection. Online (Jul 2015), http://data.europa.eu/euodp/en/data/dataset/S2075_83_1_431_ENG
14. European Union: Council regulation art. 12 regulation (eu) 2016/679 (2016)
15. European Union: Council regulation art. 15 regulation (eu) 2016/679 (2016)
16. European Union: Council regulation art. 25 regulation (eu) 2016/679 (2016)
17. European Union: Council regulation art. 4 regulation (eu) 2016/679 (2016)
18. European Union: Council regulation art. 9 regulation (eu) 2016/679 (2016)
19. Fuentes, M.R.: Cybercrime and other threats faced by the healthcare industry. Trend Micro (2017)
20. Geodakyan, G.S., Yen, Y.j.S., Foss, R.A., Hardy, J., Broen, W.D., Born, N.M.: Method and system for combining offline and online identities with associated purchasing intention indicators in view of a geographic location (Sep 18 2018), US Patent App. 15/712,036
21. Gluck, J., Schaub, F., Friedman, A., Habib, H., Sadeh, N., Cranor, L.F., Agarwal, Y.: How short is too short? implications of length and framing on the effectiveness of privacy notices. In: Symposium on Usable Privacy and Security (SOUPS). pp. 321–340 (2016)
22. Gruss, D., Schwarz, M., Wübbeling, M., Guggi, S., Malderle, T., More, S., Lipp, M.: Use-after-freemail: Generalizing the use-after-free problem and applying it to email services. In: Proceedings of the 2018 on Asia Conference on Computer and Communications Security. pp. 297–311. ACM (2018)
23. Harkous, H., Fawaz, K., Lebret, R., Schaub, F., Shin, K.G., Aberer, K.: Polisis: Automated analysis and presentation of privacy policies using deep learning. In: 27th USENIX Security Symposium (USENIX Security 18). pp. 531–548 (2018)
24. Hern, A.: Fitness tracking app Strava gives away location of secret US army bases. The Guardian **28** (2018)
25. Hertlein, M.: Digitale identitäten erfolgreich schützen. Online (Apr 2019), https://www.security-insider.de/digitale-identitaeten-erfolgreich-schuetzen-a-821563/
26. Hu, H., Wang, G.: End-to-end measurements of email spoofing attacks. In: 27th USENIX Security Symposium. pp. 1095–1112 (2018)
27. Jansen, F.: Verdächtiger nennt Ärger über Politiker als Motiv für Datenklau. Online (Jan 2019), https://m.tagesspiegel.de/politik/datendiebstahl-verdaechtiger-nennt-aerger-ueber-politiker-als-motiv-fuer-datenklau/23838452.html
28. Jensen, C., Potts, C.: Privacy policies as decision-making tools: An evaluation of online privacy notices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 471–478. CHI '04 (2004)
29. Klensin, J., Freed, N., Rose, M., Stefferud, E., Crocker, D.: Smtp service extensions. Tech. rep., RFC 2846, November (1995)
30. Linden, T., Harkous, H., Fawaz, K.: The privacy policy landscape after the gdpr. arXiv preprint arXiv:1809.08396 (2018)
31. LLC, P.I.: The race to gdpr:a study of companies in the united states & europe. Tech. rep., McDermott Will & Emery LLP (2018)
32. Martino, M.D., Robyns, P., Weyts, W., Quax, P., Lamotte, W., Andries, K.: Personal Information Leakage by Abusing the GDPR Right of Access. In: Symposium on Usable Privacy and Security (SOUPS) (2019)
33. Matwyshyn, A.M., Cui, A., Keromytis, A.D., Stolfo, S.J.: Ethics in security vulnerability research. IEEE Security Privacy **8**(2), 67–72 (2010)
34. Mell, P., Scarfone, K., Romanosky, S.: Common vulnerability scoring system. IEEE Security & Privacy **4**(6), 85–89 (2006)

35. Mueller, R.: Report on the investigation into russian interference in the 2016 presidential election. US Dept. of Justice. Washington, DC (2019)
36. Ollmann, G.: The phishing guide–understanding & preventing phishing attacks. NGS Software Insight Security Research (2004)
37. Österreich, N.: Entwurf bundesgesetz über sorgfalt und verantwortung im netz. Online (2019), https://cdn.netzpolitik.org/wp-upload/2019/04/Digitales-Vermummungsverbot-Gesetzesentwurf.pdf
38. Papageorgiou, A., Strigkos, M., Politou, E., Alepis, E., Solanas, A., Patsakis, C.: Security and privacy analysis of mobile health applications: the alarming state of practice. IEEE Access **6**, 9390–9403 (2018)
39. Pascual, A., Marchini, K.: 2018 child identity fraud study. Online (Apr 2018), https://www.javelinstrategy.com/coverage-area/2018-child-identity-fraud-study
40. Pollach, I.: What's wrong with online privacy policies? Commun. ACM **50**(9), 103–108 (Sep 2007)
41. Proofpoint: Social media brand protection fraud. Online (2017), https://www.proofpoint.com/sites/default/files/pfpt-en-social-media-protection-brand-fraud-report.pdf
42. Protenus: 2017 breach barometer annual report. Online (2017), https://www.protenus.com/2017-breach-barometer-annual-report
43. Rasthofer, S., Huber, S., Arzt, S.: All your family secrets belong to us worrisome security issues in tracker apps. In: DEF CON 26 (2018)
44. Roesner, F., Kohno, T., Wetherall, D.: Detecting and defending against third-party tracking on the web. In: Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI). pp. 155–168 (2012)
45. Schneier, B.: Doxing as an attack. Online (Jan 2015), https://www.schneier.com/blog/archives/2015/01/doxing_as_an_at.html
46. Secureworks Counter Threat Unit Threat Intelligence: Threat Group 4127 Targets Hillary Clinton Presidential Campaign. Online (Jun 2016), https://www.secureworks.com/research/threat-group-4127-targets-hillary-clinton-presidential-campaign
47. Seymour, J., Aqil, A.: Your voice is my passport. In: DEF Con 26 (2018)
48. Snyder, P., Doerfler, P., Kanich, C., McCoy, D.: Fifteen minutes of unwanted fame: Detecting and characterizing doxing. In: Internet Measurement Conference. pp. 432–444. IMC'17 (2017)
49. Snyder, P., Doerfler, P., Kanich, C., McCoy, D.: Fifteen minutes of unwanted fame: Detecting and characterizing doxing. In: Proceedings of the 2017 Internet Measurement Conference. pp. 432–444. ACM (2017)
50. Times, N.Y.: An old swindle revived. Print (Mar 1898), https://www.nytimes.com/1898/03/20/archives/an-old-swindle-revived-the-spanish-prisoner-and-buried-treasure.html
51. TrendLabs Security Intelligence: Operation Pawn Storm Ramps Up its Activities; Targets NATO, White House (2015)
52. Urban, T., Tatang, D., Degeling, M., Holz, T., Pohlmann, N.: The unwanted sharing economy: An analysis of cookie syncing and user transparency under gdpr. arXiv preprint arXiv:1811.08660 (2018)
53. Web Investigations: Investigation and Doxing Prices (Jul 2019), https://doxanybody.wordpress.com/category/investigation-and-doxing-prices/
54. Yeboah-Boateng, E.O., Amanor, P.M.: Phishing, smishing & vishing: an assessment of threats against mobile devices. Journal of Emerging Trends in Computing and Information Sciences **5**(4), 297–307 (2014)

Appendix

SAR Text in German

Betreff: Auskunft nach Artikel 15 DSGVO

Sehr geehrte Damen und Herren,
auf der Grundlage von Artikel 15 der Datenschutz-Grundverordnung (DSGVO) verlange ich hiermit Auskunft darüber, ob bei Ihnen personenbezogene Daten über mich gespeichert sind. Falls dies der Fall ist, verlange ich Auskunft über die Informationen nach Artikel 15, Absätze 1 und 2 DSGVO. Bitte stellen Sie mir die gewünschten Informationen gemäß Artikel 12 Absatz 3 DSGVO innerhalb eines Monats nach Eingang dieses Antrags zur Verfügung. Bei Nichtbeachtung meiner Forderung werde ich mich an eine Datenschutzbehörde wenden. Zudem behalte ich mir weitere rechtliche Schritte vor, die auch die Geltendmachung von Schadenersatzansprüchen nach Artikel 82 DSGVO umfassen.

Mein Name: NAME
Mein Geburtsdatum: GEBURTSDATUM
Meine gegenwärtige Anschrift:
ADRESSE

Mit freundlichen Grüßen
NAME
SIGNATURE

SAR Text translated to English

Subject: Information according to Article 15 GDPR

Dear Sir and Madam,
I hereby request, on the basis of Article 15 of the General Data Protection Regulation (GDPR), information about whether personal data about myself are stored by your company. If this is the case, I request information about the stored information based on Article 15, paragraphs 1 and 2 GDPR. Please provide me with the requested information in accordance with Article 12 paragraph 3 GDPR within one month after receipt of this subject access request. In case of failure to comply with my request, I will contact a data protection authority. I also reserve the right to take further legal action, including the assertion of claims for damages under Article 82 GDPR.

My name: NAME
My date of birth: DATE OF BIRTH
My current address:
ADDRESS

Yours sincerely,
SURNAME
SIGNATURE