Personal Information Leakage by Abusing the GDPR 'Right of Access'

Peer-reviewed author version

Personal Information Leakage by Abusing the GDPR "Right of Access"

Mariano Di Martino[1], Pieter Robyns[1], Winnie Weyts[2], Peter Quax[1,3],
Wim Lamotte[1], and Ken Andries[2,4]
[1] *Hasselt University/tUL, Expertise Centre for Digital Media*
[2] *Hasselt University - Law Faculty*
[3] *Flanders Make*
[4] *Attorney at the Brussels Bar*
{mariano.dimartino,pieter.robyns,peter.quax,wim.lamotte,ken.andries}@uhasselt.be
winnie.weyts@student.uhasselt.be

Abstract

The General Data Protection Regulation (GDPR) "Right of Access" grants (European) natural persons the right to request and access all their personal data that is being processed by a given organization. Verifying the identity of the requester is an important aspect of this process, since it is essential to prevent data leaks to unauthorized third parties (e.g. criminals). In this paper, we evaluate the verification process as implemented by 55 organizations from the domains of finances, entertainment, retail and others. To this end, we attempt to impersonate targeted individuals who have their data processed by these organizations, using only forged or publicly available information extracted from social media and alike. We show that policies and practices regarding the handling of GDPR data requests vary significantly between organizations and can often be manipulated using social engineering techniques. For 15 out of the 55 organizations, we were successfully able to impersonate a subject and obtained full access to their personal data. The leaked personal data contained a wide variety of sensitive information, including financial transactions, website visits and physical location history. Finally, we also suggest a number of practical policy improvements that can be implemented by organizations in order to minimize the risk of personal information leakage to unauthorized third parties.

1 Introduction

On the 27th of April 2016, the European Parliament and the Council of the European Union enacted Regulation 2016/679 on "the protection of natural persons with regard to the processing of personal data and on the free movement of such data" [2]. This regulation, commonly referred to as the General Data Protection Regulation (GDPR), supersedes Directive 95/46/EC and provides a number of additional benefits to natural persons (data subjects) when their data is processed by third parties (data controllers). One such example is the "Right of Access", which allows the data subject (DS) to request whether and which personal data concerning him or her is being processed by the data controller (DC) [2, Art. 15].

As of 25 May 2018, the GDPR became enforceable, meaning non-compliant DCs could face a fine of up to 20 million euros or 4% of the annual worldwide turnover of the preceding financial year, depending on the nature of the infringement [2, Art. 83]. This means that by now, DCs should have implemented the necessary controls to allow European DSs to exercise their "Right of Access" through data requests (DRs), as this right has been extended from the original Directive 95/46/EC originating from 1995. However, the modi operandi and efficacy of these controls in context of information security and privacy has, to the best of our knowledge, not been investigated in current literature. In this paper, we address exactly this issue. More concretely, we examine the following aspects of the "Right of Access":

- Which information about the DS is requested by the DC in order to verify their personal identity?
- Based on the provided information, how does the DC verify the credentials and hence the authenticity of the request?
- Can the requested information be forged by an adversary or can the DC be persuaded through social engineering such that unauthorized access to the DS's personal data is obtained?
- How can the verification of the personal identity of the DS be improved?

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.
August 11–13, 2019, Santa Clara, CA, USA.

The structure of the paper is as follows. In Section 2, we discuss the general format of a DR and how it can be used to exercise the "Right of Access". Section 3 then presents an experiment where we submitted forged DRs to 55 organizations in order to answer the research questions outlined above. Next, we propose a number of possible policy improvements for handling DRs that could be implemented by organizations in Section 4. Moreover in Sections 5, 6 and 7 we respectively discuss related work, limitations and future work, and the conclusions of this study. Finally, a more detailed discussion of the individual cases of our experiment is provided in Appendix A.1.

2 The GDPR Data Request

The "Right of Access" [2, Art. 15] introduced by the GDPR allows European consumers to request personal information from any organization that processes their data [1]. As stated in [2, Art. 4-1], "personal data" means any information relating to an identified or identifiable natural person. Practical examples of such personal data can exist of, for instance: location history, financial transactions, written messages, etc.

To exercise this right, the DS has to submit a DR to the desired organization by any means, such as email or postal mail [2, Art. 12]. As the DC should avoid leaking personal data to unauthorized adversaries, it can respond to a DR by requesting the subject to verify their identity and thus ensure that the sensitive data is delivered to the right person.

Each DC should respond to a DR with the requested information, without undue delay and in any event within one calendar month, unless an additional extension of 2 months is requested by the DC due to the complexity or the large number of current DRs [2, Art. 12.3]. This means that the subject should, in any event, at least receive a response within one calendar month and should receive the required information in no more than 3 calendar months, preferably in an electronic format [2, Rec. 59]. Furthermore, the personal data should be presented to the subject in a "commonly used electronic form" [2, Art. 15-3] and in some specific cases, also in a "structured, commonly used and machine-readable format" [2, Art. 20], meaning that – for instance – screenshots are not allowed.

In order to manage such rights effectively, a Data Protection Officer (DPO) should be appointed in organizations whose core activities consist of regular and systemic monitoring of DSs on a large scale or consist of large scale processing of sensitive data [2, Art. 37].

3 Data Request experiment

In this section, we discuss an experiment where we attempt to send unauthorized DRs by impersonating targeted individuals and therefore abuse the GDPR "Right of Access". First, we describe the assumptions from our adversarial model and lay out the communication and relations between the authors and targeted individuals in Section 3.1. Moreover, the methodology and ethical aspects on how our experiment was conducted are discussed in Section 3.2 and Section 3.3. Furthermore in Section 3.4, we analyze the different credentials that organizations request in order to verify the identity of the DS. Finally in Section 3.5, impersonation techniques are presented that can be applied to extract or forge credentials from the targeted individuals in practical scenarios.

3.1 Adversarial model

We acquired the permission to set up the experiment with 2 of our co-authors (which we will refer to as 'targeted individuals'). Our goal is to impersonate these individuals in order to obtain personal information by performing illicit DRs. First, in order to familiarize ourselves with the targeted individuals, we asked each one of them the following questions:

- The name of the targeted individual.
- A list of several (local, national or international) organizations of which they knew the organizations had personal information regarding them.
- A link to one public social media profile of the targeted individual.
- The home and email address of the targeted individual.

As we will discuss in Section 3.5.1, such information can be easily gathered from various public sources such as social media or government registers. For our two targeted individuals, we indeed found all information listed above on public sources, except for the home address. In practice, an adversarial model may be weakened or fortified depending on the relation between adversary and targeted individual.

From our targeted individuals, we collected the names of 55 unique organizations to which we posed DR as part of our study. Among these organizations, almost half of them are also present in the Belgian Alexa top 50 [3].

As described above, each of the targeted individuals has also cooperated in the composition of this study as an author of this paper. The reason for this is twofold: (1) due to our willingness to perform an ethical experiment, we were uncertain of the scope of personal data that we would receive from external volunteers when performing illicit DRs, hence minimizing an impact on privacy; (2) in a recent framework such as the GDPR, it would be useful to first analyze how different organizations handle DRs. As the DR procedure should not differ significantly between DSs, we focus on the sample size in terms of the number of organizations instead of the number of individuals we targeted.

[1] The GDPR is also applicable for EU organizations that process personal data from non-EU consumers.

3.2 Evaluation methodology

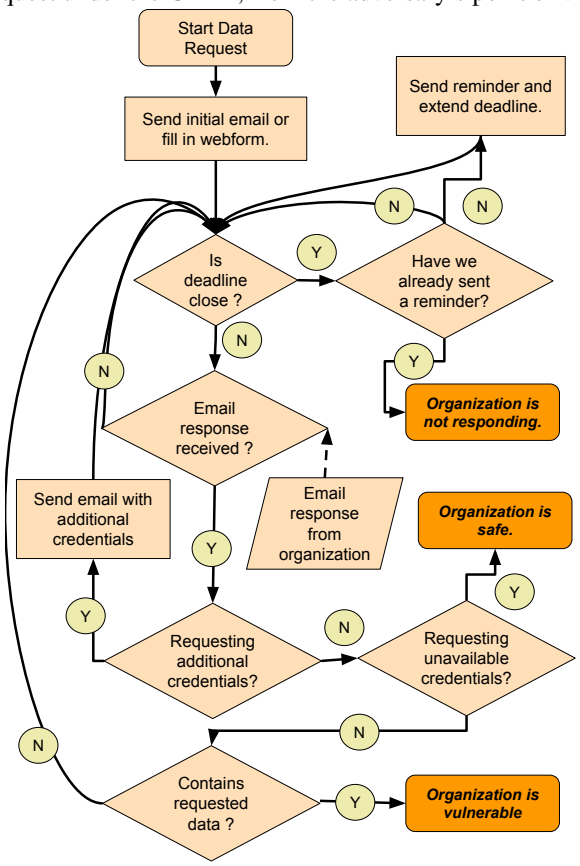
With the list of organizations from each targeted individual, we located the websites of each organization and manually extracted an email address (often located in privacy policies) or link to a web form that is provided to submit DRs. After the extraction, we created a template to exercise the "Right of Access" under the GDPR and submitted a DR to each of the organizations, either through email or by filling in the web form (which we will discuss in Section 3.5.2). With the intention to construct a credible DR, our template also included several questions regarding the retention period of personal data, automated profiling and various methods on how they collect personal information. In the remainder of the paper, the authors are henceforth represented as the adversary, while the targeted individuals are portrayed as the DS.

Our process of performing Data Requests is demonstrated in detail in Figure 1. All email communication was conducted starting from October 16th 2018 until March 12th 2019. Emails that were received on the original email address of each individual, inaccessible by the adversary, were ignored. At the end of the experiment, each organization is assigned to one of the following 3 groups:

- **"Organization is not responding"**: If the organization refrains from responding to our request after a reminder and 2 months of silence, we conclude that the organization is unwilling to fulfill our request and thus is legally not compliant to the GDPR, risking corrective actions (such as fines) [2, Art. 83] and judicial proceedings [2, Art. 79] .
- **"Organization is vulnerable"**: If the organization has delivered personal data from the targeted individual to the adversary, we then conclude that the organization is not able to correctly verify the identity of the DS. As a result, this leads to a data breach of personal information and is therefore non-compliant with the GDPR [2, Art. 88]. Consumers that utilize the services of those organizations are clearly exposed to leakages of their personal information to any determined adversary.
- **"Organization is safe"**: Organizations that do not release personal information about the targeted individual due to secure authentication mechanisms, are considered safe in the context of our adversarial model.

There are 2 exceptions to Figure 1, (1) if an organization adheres to the DR by responding to the original email address instead of the email address of the adversary, we consider this organization to be safe as long as the subject's data is not received by the adversary; (2) if the credentials requested by the organization are not available to the adversary (indicated by "Requesting unavailable credentials"), then we attempt to persuade the DC using the techniques presented in Section 3.5.4. Furthermore, deadlines of one month are

Figure 1: Our experimental process of performing a Data Request under the GDPR, from the adversary's point of view.



established unless the organization requests to extend the deadline with two months, corresponding to Article 12 [2, Art. 12]. Moreover, in case the company is considered to be safe, we assist the targeted individuals to continue the DR process in order to analyze the personal data for any incidental leaks.

On the grounds of ethical research, we do not publicly denounce organizations by name and therefore use a pseudonym that indicates the category in which the organization belongs. These categories consist of: Financial (Fin_x), Retail (Ret_x), Entertainment (Ent_x), Transport and Logistics (Trl_x), News Outlet (New_x) and Other (Oth_x) organizations.

3.3 Notes on ethical research

In compliance with ethical research guidelines, the experiment performed in this study was approved and authorized by the university Ethical Research Committee (ERC). Involved individuals were required to acknowledge, through a signed declaration, that their credentials would be used in order to submit unauthorized DRs. Moreover, the targeted individuals (co-authors) gave written permission to read any relevant email communications between them and the DCs for the duration of the experiment. Furthermore, the personal data that

we unintentionally received from the organizations regarding unrelated individuals, were immediately removed after taking note of the event. In addition, a copy of the data from the targeted individuals was sent to the rightful individuals and removed by the adversary after the experiment was finished.

Similar to a responsible disclosure model [8], all "vulnerable" organizations have been notified of the details concerning our research and were individually given advice via email on how to improve their policies of handling DRs. This interaction led to a follow-up personal meeting with the Data Protection Officers of three organizations, where the findings and suggestions for improvements were discussed more in-depth. Our approach to this study was appreciated by the DPOs, as we further ensured that the vulnerable organizations had a reasonable amount of time to implement any necessary changes to their process before publication of this study.

As we will discuss in Section 3.4.3, part of our experiment involved modifying an individual's proof of identity before sending it to an organization. It should be pointed out that no official government documents were altered during this process, only a scanned photocopy. At the same time, we acquired prior permission of the individuals whose proof of identity was used and explicitly obtained clearance from our legal council and the ERC.

Furthermore, we recognize the fact that processing a DR may lead to a certain financial cost for those organizations that handle them manually or have a significant amount of personal data about the DS. The DRs we sent out in this experiment could be considered needless and thus obtrusive to the organizations involved. To counterbalance this, we opted to contact the organizations afterwards to inform them about the outcome of the experiment and to inform them on potential improvements in their handling of DRs. This way of working was universally appreciated by all organizations involved. At the same time, it should be considered that the only way to obtain the necessary information about practical handling of DRs is by actually sending them out - these experiments cannot be performed in a confined lab context. The authors feel that the societal benefits of improving consumer privacy and the organizations' internal policy (which hopefully will be the long term outcome of this study) outweigh the financial costs.

We strongly recommend that future studies should take these ethical considerations into account when deploying such experiment on a larger scale. Finally, the considered organizations were *not* reported to third parties (e.g. the Data Protection Authority), and their identity was anonymized in this paper in order to minimize reputational damage and the risk of criminal targeting.

3.4 Authentication credentials

When a DR is submitted to a DC, the identity of the DS must be verified in order to prevent leakage of personal information to an unauthorized third party. The GDPR therefore suggests that the same authentication mechanism should be used for both DRs and for authenticating the DS to the online services offered by the DC [2, Rec. 57]. However, this practice is not explicitly enforced by law.

Recital 64 additionally states that "the controller should use all reasonable measures to verify the identity of a DS who requests access". Hence, organizations are given the freedom to choose their own policies, depending on their definition of "reasonable measures". This is corroborated further by Article 12, which states: "where the controller has reasonable doubts concerning the identity of the natural person making the request [...], the controller may request the provision of additional information necessary to confirm the identity of the DS." [2, Art. 12 (6)].

In summary, although the GDPR provides general guidelines, the precise type of information that should or should not be requested from the DS for authentication purposes is left to the discretion of the DC.[2] Over the course of the experiment we observed that in practice, organizations indeed request a wide variety of credentials to confirm the identity of their users as a result (the nature of which is typically found in the privacy statement).

In Table 3, we present an overview of all the manually contacted organizations and which credentials (authentication data) they requested in order to verify the identity of the DS. Additional details related to this table are defined as followed:

- The "Link leakage" check marks indicate whether the organization unintentionally leaked other personal information unrelated to our initial DR, which can occur in 2 cases: (1) personal data from other individuals with a similar or identical name are included in the response to a DR; (2) the organization has no email address for some account A on file, so the first account B that is created with a name and date of birth identical to account A, will be linked by the organization to account B. An adversary is able to create account B and then perform a DR for account B, resulting into a leakage of data from account A through account B.

- A check mark in the "Vulnerable" column corresponds to the "Organization is vulnerable" description, as discussed in Section 3.2.

- The column "Region" indicates the organization's market area, defined to be either "Local", "National" or "International".

The following subsections discuss the results of this table and describe the different types of required credentials that we encountered in detail.

[2] Subject to the general principles of processing personal data contained in article 5 of the GDPR, such as data minimisation.

Table 1: Number of automatic and manual DRs handling processes of organizations, including the number of answered and unanswered DRs and the number of vulnerable organizations.

DR process	Answered	Unanswered	Vulnerable
Automatic	14	N/A	0
Manual	37	4	15

3.4.1 Login credentials

In Table 1, we show that for 14 out of 55 investigated organizations, performing a Data Request is *only* possible through a dedicated web page after logging in on the organization website, as recommended in Recitals 57 and 63 of the GDPR [2, Rec. 57, 63]. An extra 3 out of the remaining 41 organizations require the DS to log in (e.g. through an external dedicated webpage of privacy management software) after the identity was verified through email communication, which is referred to as "Account verification" in Table 3.[3] In addition, one organization allowed the DS to access their personal data in multiple ways, including login credentials and another organization was persuaded to provide an alternative for the "Account verification" (shown by '*'). In summary with this type of login credentials, the DC provides only the personal data from the account associated with the credentials in question.

Observe from column "Account verification" in Table 3, that all DCs which require the user to log in are *not* vulnerable, since in these cases the DR procedure is protected by the authentication mechanism of the website. Clearly, these requirements cannot be enforced if the organization does not have a website or if data about the DS was stored without requiring the creation of an account on the organization's website. Either of these scenarios give rise to a significantly greater challenge to verify the identity of the DS, as we will discuss in Section 4.

3.4.2 Email address

Instead of requiring the user to log in, 41 organizations allow the subject to perform a Data Request by explicitly emailing the DPO or DC, whose email address is typically found in the privacy statement on the organization website. As such, the request is manually handled or at least, analyzed by a human correspondent. The DPO/DC should ideally only adhere to the request if it is made from the same email address with which the user is registered on the organization's website.

However, only 12 of 41 organizations enforced this policy and an additional 5 organizations permitted the subject to offer other credentials if the subject no longer has access to their original email account. In most cases, specific user data (e.g. last products bought) was requested to compensate for not being able to access the original email account. Considering that a realistic adversary has no access to such information, and assuming this information cannot be trivially guessed, we consider these organizations to be safe unless a link leakage has occurred such as for example in the case discussed in Appendix A.1.3.

[3] This is different compared to the "automatic" process, as such a process does not allow a DS to initially request their personal data by email.

3.4.3 National identity card

Another credential required by 13 out of 55 organizations is a digital copy or scan of the national identity (ID) card of the subject. The copy is either uploaded via a web form dedicated to DRs or included as attachment in case the DR is performed via email. One organization requested the front and back side of the ID card, while the remaining 12 organizations only requested the front side. Note that while the National Register Number (NRN) is only written on the back side, the Card Identification Number (CIN) is located on the front side of the ID card. However, since "a controller should not retain personal data for the sole purpose of being able to react to potential requests" [2, Rec. 64], sensitive data on the ID card that is known not to be in possession of the DC, e.g. NRN and CIN, can be censored by the subject [2, Art. 25]. In fact, this was explicitly required by default for 11 organizations.

3.4.4 Home address

A lesser used credential is the home address of the subject, required by 5 out of 55 organizations. Four of these request the complete address consisting of the street name and city, while the remaining organization only demands the region in which the subject lives such as the city or province. Generally speaking, knowing the region of the subject is a relatively easy task for a determined adversary given that social media accounts often disclose this information; it can also be obtained through various public databases as we will discuss in Section 3.5.1. Likewise, even the complete address of the subject might be available (although this information is typically contained in other sources).

Forms of Human Intelligence (HUMINT), where the adversary might be able to communicate directly with the subject or friends of the subject, is also a valuable approach to steal the necessary information. Phishing campaigns are clearly an effective method to extract such personal information.

3.4.5 Calling the subject

Calling the subject on a phone number known by the DC beforehand is a safer authentication method, but is unfortunately only carried out by 2 out of 55 organizations. By making a call, the DC can speak directly to the DS and as such confirm the submission of a DR or request additional user-specific data for the purpose of authentication (see for example Appendix A.1.2).

For an adversary, intercepting calls to the DS's phone is difficult, although possible through for example additional so-

cial engineering [6]. On the other hand, spoofing the caller ID of the subject is a relatively trivial task [11], but has no useful purpose in this scenario as the DC calls the subject and not the other way around. In case the subject performs the initial DR orally (for instance; through a phone call), the DC must still verify the identity through other means [2, Art. 12.1], presumably to avoid precisely such an identity spoofing attack.

As we had no access to the mobile phone of the targeted individuals, we concluded that organizations that performed this authentication method are safe in the context of our adversarial model.

3.4.6 Specific user data

The final credential that we discuss is a demand of the DC to provide specific user data from the DS, requested by 11 out of 55 organizations. This includes various unrelated pieces of information, depending on the nature of the organization. For instance, an entertainment venue might ask to provide the date of the last visit and the products that were bought by the DS.

Determining this information for an adversary is challenging and usually requires in-depth knowledge of the DS. Here, Open Source Intelligence (OSINT) methodologies are useful to e.g. find photos that indicate visits but are in many cases not sufficient to discover the exact details required. Due to difficulty of extracting the necessary information, we consider organizations that request such specific data to be safe in the context of our adversarial model.

3.5 Impersonation techniques

In order for an adversary to obtain the subject's personal data through a DR, they must trick the DC into believing the request is legitimate by impersonating the subject. Due to the non-explicit nature of Recital 64, there is no one-size-fits-all approach to achieve this goal (which is also true for social engineering in general), and a determined adversary is more likely to devise an impersonation strategy that is specifically tailored to meet the set of requirements mandated by one specific organization. In this section, we discuss the impersonation techniques useful in forging or extracting the necessary credentials.

3.5.1 Intelligence gathering

As impersonation strategies often demand information from external sources, we explore a number of different intelligence techniques that are able to fabricate a trustworthy profile of our targeted individual. In this section, we merely explain the possible methods of extracting basic information useful to perform illicit DRs.

The most common approach is Open Source Intelligence (OSINT), a form of collecting publicly available information from the targeted individual. Especially in society today, social media plays an important role in extracting personal data. Unsurprisingly, 79% of all people that have Internet access are in possession of at least one social media account [16]. Various social media platforms have different pieces of sensitive information depending on how strong an individual has chosen to shield that information.

For instance, a basic public version of a social media profile often consists of numerous personal images that could be used to alter a photo of an identity card. In cases where the adversary is able to open up the profile by either requesting to become friends or following the targeted individual, sensitive information becomes much more accessible. For example, the date of birth or the region of residence often becomes visible, which is information essential to employ impersonation strategies. In some extreme cases, images of purchase deeds or result sheets of driving examinations are uploaded which clearly display the address of the targeted individual. Additional leakages are also possible by discovering matches between what people like or analyzing the social media profiles of relatives [22]. Besides social media platforms, central government agencies such as the NBB (National Bank of Belgium) or telephone directories such as De Witte Gids also contain personal information (often publicly accessible).[4] Another possibility is to utilize global OSINT search engines such as Pipl [25], which permit adversaries to collect a significant amount of data from individuals with minimal effort.

As opposed to OSINT, a more rigorous and tedious approach called HUMINT is also viable to extract sensitive information from a targeted user. HUMINT serves as the basis for phishing campaigns, in which unsuspecting victims are contacted and then tricked into releasing personal identifiable data by using social engineering techniques [21]. In the context of our proposed impersonation strategies, only weak phishing campaigns are necessary where the targeted individual is able to provide us the personal information we require. However, not only the Internet is a profitable source for personal information; television and public appearances may also increase the risks of extracting valuable intelligence related to e.g. public figures.

Another source of information available to the adversary could stem from a possible personal relation with the targeted individual. For instance, a spouse may already have a significant amount of information available and therefore, would not be required to perform any lookups on social media. In fact, close relatives that reside in the same household such as a spouse, brother or sister might even be able to access the smartphone of the individual, thereby circumventing the "Call subject" authentication method. To the contrary, a person unknown to the targeted individual may not have access to the physical address and therefore has to consult additional sources to collect this information.

To conclude, we argue that excerpting enough personal

[4] "De Witte Gids" (https://dewittegids.be) is the Belgian version of a "White Pages" directory.

identifiable information from a socially active user is feasible, given the many possibilities for a determined adversary.

3.5.2 Email address spoofing

A common and basic strategy to impersonate a user (subject) is to spoof their registered email address, which we will henceforth refer to as the *original email address*. Any email address controlled by the adversary will be denoted as a *fake email address*. In our experiment, we applied a number of techniques to impersonate the targeted individuals via email:

- **"The Reply-To"**: The adversary sets the `From` header of the email to the original email address and the `Reply-To` header to a fake email address before sending. Upon replying to this email, email clients should automatically fill in the email address from the `Reply-To` header as the destination.[5] Furthermore, at the time of writing, most popular email clients (for example Gmail and Outlook), only show the `From`, `To`, and `CC` fields to the user when an email is opened, whereas the `Reply-To` field is hidden by default. As a result, an inattentive handler of the DR could be tricked into thinking that the DR originated from a legitimate user, while the reply is sent towards an email address under control of the adversary.
- **"The Resembler"**: The adversary registers a domain name that is similar to the domain of the original email address by using homographs. This is similar to the homograph attack described in the work of Gabrilovich and Gontmakher [12], except that the letters need to be in the same script as per ICANN guidelines [18, p. 2]. The DR is then sent from a fake email address on this domain.
- **"The Ringer"**: The adversary creates a fake email address that is identical to the original email address except for the domain, and sends the DR using this email. For example, if the original email address is "john.doe@gmail.com", the adversary will send the DR using "john.doe@protonmail.com".

Although in our experiment we only employ these techniques exactly as described above, it should be noted that in practice, many variations could be improvised. As an example, consider the case where an adversary uses "The Resembler" technique to submit a DR through a spoofed email, except this time they do not register the homographic domain name. This will render the organization unable to respond to the DR, as the domain name is not registered. Next, after a certain period of time (for example 30 days), the adversary sends a reminder email from a different email address under their control, which cites the first DR request that was transmitted with the spoofed email address. Upon receiving the reminder, the DC may recall that they were indeed unable to reply to the first DR, and be inclined to respond to the reminder email. This is exacerbated by the fact that the citation of the original email may give a false sense of legitimacy to the reminder email, despite that it was sent from a different email address under control of the adversary.

[5] It should be noted that RFC 2822 does not explicitly require that replies must be sent to the `Reply-To` address [26].

Table 2: Brief experiment to choose the best impersonation strategy by sending a DR to 15 organizations (5 per technique) and count the received responses to the adversary email address.

Technique	Received	Not received
The Reply-To	1	4
The Resembler	4	1
The Ringer	5	0

Continuing our study, the question now remains which impersonation strategy should be chosen by the adversary and how much information should be included in the original DR in order to maximize the probability of success. For finding the best email spoofing technique from the techniques discussed above, we performed a brief experiment involving 15 organizations, where each of the email spoofing techniques was used to contact 5 organizations. The results of this experiment are shown in Table 2.

As shown in the table, the "Ringer" technique resulted in the highest probability of receiving a reply to the adversary's email address, whereas the other techniques were less successful. This may be attributed to a number of reasons: the "Reply-To" technique fails if the DR email is forwarded to another person, in which case the Reply-To header is dropped. Similarly, the header may be dropped if the organization uses a ticketing system for handling emails. In these cases, the replies to the DRs were sent to the original email address instead of the fake one. Another disadvantage of the "Reply-To" technique is that it cannot be used if the organization uses a web form to submit DRs.

For testing the "Resembler" technique, one could attempt to register the domain name `protonmail.com` (Cyrillic `a`), which is similar to the domain `protonmail.com` of an account owned by one of the targeted individuals. However, this approach would fail because registering mixed-script domain names is disallowed by ICANN [18, p. 2] for the purpose of countering homograph attacks. Instead, we registered the domain name `protonmaíl.com` (note the accented 'i'), which contains letters that all belong to Latin script. Similarly to the "Reply-To" case, we noticed that some replies to the Data Requests were sent to the targeted individual's email address. If the email is handled manually by a customer service representative, this may occur if the reply's destination email address is typed manually or if it is corrected by the representative. Moreover, if the organization uses a web form, the DR was in some cases rejected altogether because of the invalid character í in the domain name.

Figure 2: A "John Doe" example of our altered ID card. Metadata of the PNG file such as the image dimension was also modified to increase the credibility of it being captured by a real photo camera.

The email spoofing types used for each organization are abbreviated in Table 3 as "Res", "Rin" and "Rep" for respectively The Resembler, The Ringer and The Reply-To.

3.5.3 Identity card image manipulation

Recall from Section 3.4.3 that if a photo or scan of the front side of an ID card is requested as an authentication credential, sensitive information such as the CIN, hand written signature and validity date number can be censored. Any information that could technically be used as a unique authentication credential that is unknown to the adversary is thereby removed. Consequently, for an adversary to successfully alter a digital copy of an ID card, only the subject's name, photo, and birth date must be known. Though physical ID cards are designed to be difficult to fabricate, a digital copy can be trivially falsified using image manipulation software as depicted in Figure 2.

In this experiment, we replaced the name, birth date, and photo on a reference ID card image to the credentials of the targeted individuals so as to create a manipulated ID card image. The targeted individuals' credentials (photo and date of birth included) were obtained through OSINT from one of their social media accounts. Using the altered ID card image, we were able to successfully authenticate as the targeted individuals in 7 out of 13 organizations that requested the ID card as part of the DR. The remaining 6 organizations requested, in combination with an ID card, additional credentials which we were not able to forge.

Despite having used a legitimate photo on the ID card of each targeted individual, it is unclear whether organizations that have a photo of the subject on file, actually compared them. If not, a stock photo could have been used, thereby reducing the number of known credentials even further and simplifying the process of creating an altered ID.

Clearly, a digital image of an ID card is not an optimal credential for authenticating users. Ignoring the privacy risks, even if the DC would ask for an uncensored NRN and be able to verify it, using the ID card as a credential would still be insecure: if leaked, NRN cannot be changed. Unfortunately, such events have occurred in the past before [5, 9].

3.5.4 Social engineering

Even when a DC requests credentials that are unavailable to the adversary in order to verify the identity of the DS, we found that in practice, the handler of the DR can sometimes be persuaded to offer alternative verification methods through social engineering. The success rate of this approach depends on various factors, including: the personality and current mood of the DR handler [28], the flexibility of policies implemented by the organization and whether employees are trained to recognize social engineering attempts [15].

Specifically in our study, we were able to persuade 8 out of the 41 DCs that handle DRs manually to diverge from standard procedure. In general, we employed the following strategies to attempt to persuade the DC:

Dismissing access to the DS's email address: When the DC requires that the DS's registered email address must be used to request or to receive personal data, the adversary can attempt to avoid this requirement by stating that they "no longer have access to this email address". For Ent_A, Ent_D, and New_B, an alternative verification method was offered where the adversary was asked to provide specific user data (which they do not have). These organizations are therefore not vulnerable. Ret_B allowed the adversary to provide an ID card as an alternative, but always required the user to log in to actually download their personal data. Trl_C on the other hand sent the DS's data to the adversary's email address without any additional verification.

Dismissing access to the DS's online account: If the DC sends the requested data via the online platform of the organization, as is the case for Fin_C, the adversary cannot retrieve the requested data. In such a scenario, the adversary can pretend that the requested data was never delivered by sending a reminder email. Fin_C responded to this by sending the requested data again via postal mail to the DS. Although the adversary also cannot intercept the DS's postal mail, the established trust with the DC allowed the adversary to request for the rectification of personal data (see Appendix A.1.1 for details of this interaction). Interestingly, Fin_C's online platform implements a two-factor authentication mechanism for logging in, and as such the adversary essentially managed to bypass this mechanism by performing a DR.

Deliberately omitting unknown credentials: The DC Fin_A by default requires the DS to provide both the front and back side of their ID card. Because the back side of a Belgian ID card contains the NRN, which the adversary does not know, the adversary requested to omit this information "due to pri-

vacy concerns". More specifically, Ent_L required a product serial number, name and date of birth of the DS, of which the adversary simply omitted the product serial number without further explanation. Since the GDPR is not explicit in stating which credentials are sufficient [2, Rec. 64], we postulate that Fin_A and Ent_L agreed to provide the personal data to the adversary in light of maintaining positive customer relations.

Naturally, social engineering is only possible if the organization allows the adversary to interact with a person at some point during the DR handling process. Therefore, the risk of successful persuasion through social engineering can be mitigated by implementing an automated DR handling process that can be initiated by the DS upon successfully authenticating on the organization website, as described in Section 3.4.1.

3.6 Types of personal data leakage

In the previous sections, we outlined the various credentials requested by organizations in order to verify the identity of the DS and demonstrated how an adversary can use impersonation techniques to pass the verification process in the interest of obtaining personal data of a targeted individual. We will now present an overview of the various types of personal information that were leaked by the organizations considered in our study. Since listing the types of data leaked by each organization individually could reveal the identity of the organization, we group the personal information leakages per organization category:

- **Financial institutions**: ID card number, list of timestamped financial transactions, customer ID, telephone numbers, place of birth, partial debit and credit card numbers, list of products purchased from the financial institution, and account numbers.
- **Retail**: List of purchased products, information on purchased products (e.g. serial number), sold products, and delivery dates.
- **Entertainment**: Purchased products and preferences.
- **Transport and logistics**: Timestamped visited locations with GPS coordinates, saved routes, purchased tickets, purchased subscriptions, and customer ID.
- **News outlets**: Browsing history, personal preferences and information about the device used to visit the news outlet's website (e.g. browser and operating system).

Aside from the personal data listed above, each category also leaked the full name, home address and email address of the targeted individual. The data was delivered to the adversary via email as either a `pdf`, `csv`, `xls`, `doc`, text or screenshot attachment. Note that the information obtained from each of these organizations could in practice be "daisy chained" to increase the credibility of DRs to other organizations, although we did not consider this type of adversary in our study.

3.7 Summary of results

As shown in Table 1, we have analyzed the policies of 55 organizations, of which 14 have an automated process and 41 process requests manually. From the latter, there were 4 organizations that did not respond to our DR, even after repeated attempts.

None of the organizations with an automatic process had any "Link leakage". However, 15 out of the 41 manually contacted organizations have leaked personal data from the targeted individual to an unauthorized third party. Ignoring the organizations with a "Link leakage", there are still 12 organizations that are left vulnerable to illicit DRs.

Interestingly, financial organizations – which should have a higher responsibility and higher standard of compliance required to safeguard personal information – are vulnerable in 4 out of 5 considered organizations, as shown in Table 3. To the contrary, only 2 out of 12 considered entertainment organizations are vulnerable. From the total of 15 vulnerable organizations, there are 8 organizations which would not have been vulnerable without an altered ID card. Meanwhile, the remaining 7 organizations were exploitable by persuading the DR handler or by using extracted information from OSINT sources.

4 Improving Data Request authentication

Based on the findings of our study, we propose several recommendations for organizations on how to securely handle a DR and for consumers on how to protect themselves against identity theft in the context of DRs.

4.1 Recommendations for organizations

Our results have shown that a substantial number of existing GDPR policies that implement authentication methods for DRs are clearly inadequate. Nevertheless, Recital 57 of the GDPR [2, Rec. 57] suggests DCs to verify the identity of the subject by offering a dedicated service where a subject is able to authenticate him/herself by providing the same credentials used for the online platform of the DC. From a technical viewpoint, we agree that the suggestion in the current recital is an effective method, as there would be no increase in risk resulting from having a separate authentication mechanism specifically for handling DRs. Due to the automated nature of such a service, it also minimizes the risk of link leakages and social engineering.

Despite this being a useful method, small to medium-scale organizations usually do not have the resources to realize such a service as it often requires expensive architectural changes in order to build them in a secure and reliable way [23]. In case an organization is unable to build the aforementioned service but still has knowledge of an email address of the subject, we suggest the DC to strictly adhere to a policy of accepting DRs only from precisely this registered email

Table 3: Overview of requested credentials and the resulting susceptibility for leakages from all 37 organizations that responded to our manual DR. The columns denote the required credential, while the rows indicate the pseudonyms of each considered organization. An asterisk shows that the corresponding credential was not forced, by either accepting an alternative credential or by being able to persuade the DC (*).

Organization	Account verification	Access to user email	Date of birth	Region of residence	Address	Front ID	Back ID	Call subject	Specific user data	Link leakage	Vulnerable	Region	Spoofing type
Fin_A			✓			✓	*				✓	N	Rin
Fin_B			✓								✓	N	Res
Fin_C	*		✓			✓					✓	I	Rin
Fin_D			✓			✓					✓	I	Rin
Fin_E		✓	✓					✓	✓			I	Rin
Ret_A			✓								✓	L	Rin
Ret_B	✓	*	✓			*						I	Rin
Ret_C		✓	✓									I	Rin
Ret_D			✓								✓	N	Res
Ret_E				✓							✓	N	Res
Ret_F			✓	✓	✓	✓					✓	I	Rin
Ret_G									✓			N	Rep
Ret_H		✓	✓									I	Rin
Ret_I			✓							✓	✓	N	Rin
Ent_A		*							*			I	Rin
Ent_B		✓	✓	✓	✓	✓						N	Rin
Ent_C									✓			I	Rin
Ent_D		*	✓						*			I	Rin
Ent_E		✓										N	Rep
Ent_F			✓						✓			I	Rin
Ent_G			✓								✓	L	Rep
Ent_H		✓	✓						✓			I	Rin
Ent_I	✓	✓	✓			✓						I	Rin
Ent_J		✓	✓									N	Rep
Ent_K		✓	✓									N	Rep
Ent_L			✓						*		✓	I	Rin
Trl_A		✓	✓	✓	✓					✓	✓	N	Res
Trl_B		✓	✓			✓						N	Res
Trl_C		*	✓			✓					✓	I	Rin
Trl_D			✓			✓				✓	✓	N	Rin
New_A			✓	✓	✓	✓					✓	N	Res
New_B		*	✓			✓			*			N	Rin
Oth_A	✓											N	Res
Oth_B			✓			✓		✓				N	Rin
Oth_C									✓			I	Rin
Oth_D		✓	✓									I	Rin
Oth_E									✓			N	Rin

address. Interestingly, some DCs such as the one discussed in Appendix A.1.1 had knowledge of the original email address (as the email address was contained in the data package) but nevertheless did not mandate this policy. However – even if such policy is adhered to – an adversary that has access to the mailbox of the subject, might still be able to bypass any two-factor authentication (which is potentially required when attempting to log in to the service by normal means).

A more concerning issue is the fact that some DCs are not in possession of online credentials (e.g. email addresses and passwords), making it impossible to implement such a policy. In case the organization does however have the phone number of the subject, we propose to call the subject to verify their identity, even though it requires a human operator, which might be an even greater burden on small scale organizations [14].

A final authentication method that we consider is to request user-specific data from the subject. For instance, an electricity company might ask for multiple reference numbers located on one of the subject's invoices, while an insurance company is able to request similar information located on insurance papers. However, care needs to be taken as some user specific data might still be easy to deduce, depending on the type of the organization.

Nonetheless, there are situations where the DC has no useful information to verify the identity of the DS. In these cases, the "Right of Access" does not apply and hence, the DS is unable to perform a DR to that organization, unless the DS provides additional information that enables the DC to verify the identity [2, Art.11]. Moreover, recital [2, Rec. 57] suggests DCs to not retain information that is "for the sole purpose of complying with any provision of this Regulation.". In other words, the DC should not retain personal information from the DS with the *only* purpose to respond to possible DRs, thereby significantly reducing the number of available authentication methods. As a result, it is difficult to propose an authentication method in case the DC has an insufficient amount of information to verify against.

A summary of the authentication methods that we propose are listed below, in decreasing order of importance regarding privacy and viability:

1. An automated process that requires known login credentials, without the ability to bypass an existing 2-factor authentication such as SMS messages.
2. A *strict* policy of only permitting DRs for online accounts that are sent by the email attributed to that and only that account. In addition, call the subject and request specific user data.
3. Call the subject and request specific user data.
4. Request specific user data.

It is evident that proposing a one-size-fits-all approach is problematic. Commercial tools that aid in processing DRs do exist, but it is unclear how reliable and effective those tools are.[6] In conclusion, we also suggest for employees that handle such requests to be trained in how to detect and avoid impersonation strategies in order to securely process DR.

4.2 Recommendations for consumers

Regardless of the fact that organizations are primarily responsible for providing improvements, we also present several options for DSs to reduce the chances for future data breaches. As social media is currently used by approximately 45% of all people in the world [16], sharing more personal data poses a significant risk for identity theft in general. Even though removing social media profiles entirely would substantially reduce the risk of illicit DRs, it is often an unrealistic suggestion for many consumers. More realistically, personal information such as profile photos and posts should be hidden from the public and only accessible for (close) friends, thereby shrinking the set of available data to possible adversaries. Many platforms allow consumers to fine-tune their privacy settings separately for each piece of personal information [27]. Nevertheless, we recommend users to completely hide sensitive information that could reveal their date of birth or region of residence on social media platforms as this information may be utilized by adversaries to devise a credible DR.

In addition, consumers should be attentive to emails that contain information related to DRs as they might disclose possible impersonation attempts by adversaries. For instance in Appendix A.1.1, the organization first sends the personal data to the legitimate DS on the online platform, thus indirectly notifying the DS of a failed DR attempt. In this unfortunate event, we recommend consumers to take preliminary measures by contacting the organization in question such that potential data breaches can be mitigated.

As a last recommendation, we suggest consumers to think carefully about the services or products they buy from the corresponding organizations. A quick look at the privacy policy of a given organization might already provide a rough judgement about the importance of privacy in that organization. Furthermore, performing a legitimate DR as a consumer will divulge most of the credentials necessary in the organization's process of verifying the DS's identity. Clearly, requesting credentials such as an ID card or basic personal information may indicate a poor GDPR policy for handling DRs.

5 Related work

To the best of our knowledge, Galetta et al. [13] were the first to empirically examine the practicality of performing DRs in Belgium under the now repealed Directive 95/46/ EC [1]. In their work, they showed that DCs were often insufficiently prepared to handle such requests as only 11 out of 19

[6] e.g. OneTrust and Jumio

organizations responded to their initial DR. Two years later, Ausloos et al. [4] confirmed the lack of privacy awareness with another 60 services and further discuss the difficulties that DSs encounter when attempting to exercise their rights.

After the GDPR went into effect, Wong et al. [29] exercised several consumer rights introduced by the GDPR with 230 different organizations and showed improvements in terms of usability compared to previous work, but demonstrated inadequacy in data formats. Furthermore, the authors briefly touched upon the various authentication methods that were required by the DCs in which only 88 out of 230 organizations required additional credentials. Surprisingly, 62 out of 230 DCs did not provide the subject with the personal information that was mandated after a period of 3 months. However, their experiment had a different approach compared to ours as they did not attempt to impersonate other DSs and furthermore, did not discuss the different credentials required in the point of view of an unauthorized adversary.

More recently in 2019, additional studies regarding the "Right of Access" have been conducted to show the negligent behaviour of organizations as some of them still do not correctly adhere to the subjects' rights or flat-out refuse to handle the DRs [7, 10, 24].

Though the "Right of Access" has not been subject to social engineering techniques in related work, there are a number of works that explore such techniques in an OSINT context [17, 20].

6 Limitations and future work

Our study has a number of limitations that could be addressed in future work. First, the set of targeted individuals is limited in size, as it consists of two co-authors. Even so, we argue that this limitation does not discredit our findings as an organization's DR handling *process* ideally *should* not differ from one DS to the other. We also postulate that recruiting a large number of participants for similar studies will prove to be a difficult task, since there is a significant risk involved for each participant. Indeed, as mentioned in Section 3.6, a large quantity of highly sensitive information about the participant may leak. The participants must fully trust that such leaks will not be abused by the researchers. If so, considering a larger set of targeted individuals in a future study, with multiple DRs per organization, would reduce the probability of false negatives (organizations that have a poor policy but where the adversary got "unlucky" and information was not leaked). Furthermore, biases towards certain ethnicities, professions or nationalities could be identified. It should however be noted that, with an increased number of DRs directed towards a single organization, additional care must be taken not to raise suspicion.

Second, our study considered 55 organizations, coming from a broad range of industries. Although we believe this ensures the generalizability of our findings, it might be interesting for future studies to focus more on one specific industry in order to discover any characteristic patterns pertaining to that industry.

A final limitation is that our study cannot precisely define the required credentials for a successful DR. This is due to the fact that DR handling processes differ significantly between organizations and are not fully disclosed in a public way. Furthermore, for organizations that make human interaction part of the process, the success of a DR is also dependent on the personality of the DR handler. Subsequent studies may therefore consider the rigorousness of an organization's policies and how those can be transferred and abused in related rights such as the "Right to Rectification" [2, Art. 16].

7 Conclusion

In this paper, we have explored the different credentials (authentication methods) that are requested by organizations in order to verify the identity of DSs under the "Right of Access". Additionally, different social engineering techniques have been applied to realistically forge these required credentials.

As a result, we have demonstrated that a significant number of policies for handling GDPR DRs are vulnerable due to either weak authentication mechanisms or the involvement of humans to carry out the processing of the DRs . Out of 55 examined organizations, 15 have leaked sensitive and personal information from the targeted individuals participating in our experiment, including but not limited to financial transactions, website visit histories and timestamped locations. Exercising the "Right of Access" while impersonating a DS is therefore an appealing attack for criminal adversaries.

Furthermore, we have proposed well-established authentication methods to improve the DR policy within the current legal framework. Yet, as some organizations are unable to perform these proposed methods due to not being in possession of the appropriate authentication credentials, we acknowledge that these organizations still run an increased risk of unintentionally leaking personal data to a determined adversary. We conclude that precautions also have to be taken by consumers, as it is possible to obtain valuable information through OSINT, which – as we have shown – might ultimately lead to a substantial impact on privacy.

Acknowledgements

This research was funded in part by the Bijzonder Onderzoeksfonds (BOF) of Hasselt University and by a Ph.D. Grant of the Research Foundation Flanders (FWO), grant number 1S14916N. Finally, we thank the reviewers and shepherd for their in-depth feedback.

References

[1] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data). *OJ L 281* (November 1995), 31–50.

[2] 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). *OJ L 119* (May 2016), 1–88.

[3] ALEXA. Top Sites in Belgium. https://www.alexa.com/topsites/countries/BE, accessed on January 25th 2019.

[4] AUSLOOS, J., AND DEWITTE, P. Shattering one-way mirrors - data subject access rights in practice. *International Data Privacy Law 8*, 1 (03 2018), 4–28.

[5] BARRETT, B. Hack Brief: An Astonishing 773 Million Records Exposed in Monster Breach. https://www.wired.com/story/collection-one-breach-email-accounts-passwords/, accessed on January 27th 2019.

[6] BARRETT, B. How to Protect Yourself Against a SIM Swap Attack. https://www.wired.com/story/sim-swap-attack-defend-phone/, accessed on January 15th 2019.

[7] BONIFACE, C., FOUAD, I., BIELOVA, N., LAURADOUX, C., AND SANTOS, C. Security analysis of subject access request procedures how to authenticate data subjects safely when they request for their data. In *Annual Privacy Forum* (2019).

[8] BUGCROWD. What is Responsible Disclosure? https://www.bugcrowd.com/resource/what-is-responsible-disclosure/, accessed on May 17th 2019.

[9] DAVID VOLODZKO. Marriott Breach Exposes Far More Than Just Data. https://www.forbes.com/sites/davidvolodzko/2018/12/04/marriott-breach-exposes-far-more-than-just-data/, accessed on January 15th 2019.

[10] DEMEYER, S., AND VANRENTERGHEM, A. Wat weten bedrijven echt van u? Het blijft vaak onduidelijk (Dutch). https://www.vrt.be/vrtnws/nl/2019/01/28/privacyonderzoek/, accessed on February 21st 2019.

[11] FEDERAL COMMUNICATIONS COMMISSION. Caller ID Spoofing. https://www.fcc.gov/consumers/guides/spoofing-and-caller-id, accessed on January 25th 2019.

[12] GABRILOVICH, E., AND GONTMAKHER, A. The Homograph Attack. *Commun. ACM 45*, 2 (Feb. 2002), 128–.

[13] GALETTA, A., FONIO, C., AND CERESA, A. Nothing is as it seems. the exercise of access rights in Italy and Belgium: dispelling fallacies in the legal reasoning from the 'law in theory' to the 'law in practice'. *International Data Privacy Law 6* (11 2015), ipv026.

[14] GDPR REPORT. GDPR is being abused by cyber-criminals to breach complacent businesses. https://gdpr.report/news/2018/07/04/gdpr-is-being-abused-by-cyber-criminals-to-breach-complacent-businesses/, accessed on February 10th 2019.

[15] GRAGG, D. A multi-level defense against social engineering. *SANS Reading Room, March 13* (2003).

[16] HOOTSUITE. Global Digital Report 2019. https://hootsuite.com/pages/digital-in-2019.

[17] HUBER, M., KOWALSKI, S., NOHLBERG, M., AND TJOA, S. Towards automating social engineering using social networking sites. In *2009 International Conference on Computational Science and Engineering* (Aug 2009), vol. 3, pp. 117–124.

[18] ICANN. Guidelines for the implementation of internationalized domain names. https://www.icann.org/en/system/files/files/idn-guidelines-02sep11-en.pdf, accessed on January 15th 2019.

[19] INGBER, S. Amazon Customer Receives 1,700 Audio Files Of A Stranger Who Used Alexa . https://www.npr.org/2018/12/20/678631013/amazon-customer-receives-1-700-audio-files-of-a-stranger-who-used-alexa?t=1549965015007, accessed on February 10th 2019.

[20] IRANI, D., BALDUZZI, M., BALZAROTTI, D., KIRDA, E., AND PU, C. Reverse social engineering attacks in online social networks. In *Detection of Intrusions and Malware, and Vulnerability Assessment* (Berlin, Heidelberg, 2011), T. Holz and H. Bos, Eds., Springer Berlin Heidelberg, pp. 55–74.

[21] KROMBHOLZ, K., HOBEL, H., HUBER, M., AND WEIPPL, E. Advanced social engineering attacks. *Journal of Information Security and applications 22* (2015), 113–122.

[22] LAM, I.-F., CHEN, K.-T., AND CHEN, L.-J. Involuntary information leakage in social network services. In *Proceedings of IWSEC 2008* (2008).

[23] LEONID BERSHIDSKY. Europe's Privacy Rules Are Having Unintended Consequences. https://www.bloomberg.com/opinion/articles/2018-11-14/facebook-and-google-aren-t-hurt-by-gdpr-but-smaller-firms-are, accessed on January 25th 2019.

[24] NOYBD. Netflix, spotify & youtube: Eight strategic complaints filed on "right to access". https://noyb.eu/access_streaming, accessed on January 27th 2019.

[25] PIPL. Pipl. https://www.pipl.com/, accessed on January 25th 2019.

[26] RESNICK, P. RFC 2822: Internet Message Format. *Qualcomm Incorporated* (2001).

[27] STAY SAFE ONLINE. Manage your privacy settings. *NCSA* (2019). https://staysafeonline.org/stay-safe-online/managing-your-privacy/manage-privacy-settings/.

[28] UEBELACKER, S., AND QUIEL, S. The social engineering personality framework.

[29] WONG, J., AND HENDERSON, T. How portable is portable?: Exercising the GDPR's right to data portability. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (New York, NY, USA, 2018), UbiComp '18, ACM, pp. 911–920.

A Appendix

A.1 Discussion of individual cases

As this paper generally only provides a statistical overview of the authentication methods and their subsequent breaches, we would like to present a few examples of email communication between the adversary and DC.[7] In the following sections, we will use *"my"* to describe the possession of the targeted individual while acting as an adversary. Moreover, we indicate the email from the targeted individual as the "original email" and identify the "DS" as the targeted individual.

A.1.1 International financial institute: Fin_C

The privacy policy of *Fin_C* (DC) states that the front of the subjects' identity card is required to submit a valid DR. As we submitted our request with a successful "Ringer" strategy, all electronic communication (responses included) was established with the fake email address. Moreover, the initial DR contained the name, date of birth and identity card of the targeted individual. An automatic confirmation email was received shortly after. This time-sheet depicts the subsequent communication between the adversary and the DC:

20/11/2018: Automatic email confirming the reception of our DR.

6/12/2018: The data containing all personal information was received on the online platform of *Fin_C*, which is virtually impossible for an adversary to access as it requires logging into the targeted individuals' account.

15/12/2018: In conformity with our process methodology, we sent a persuasive reminder to notify the DC that the legal deadline for responding to a DR is closing in and has a remaining 5 days left.

17/12/2018: An email was received from the DC, justifying that the data have already been sent to the aforementioned account. In addition, the DC proposed to provide a "copy of my data".

17/12/2018: We responded that we did not receive such data on "my" account and agreed to accept the "copy".

18/12/2018: The DC confirmed to send a "copy of my data".

21/12/2018: A physical copy from the data was received on the targeted individuals' home address.[8]

23/12/2018: Even though, the adversary is not aware of the specific contents of the personal information, they have however the ability to know the type of information that is provided by the DC by issuing a legitimate DR. Therefore, we sent the controller a request (still with the adversary email address) to modify "my" personal information as depicted in [2, Art. 68]. More specifically, we demanded to remove the phone number and modify the education degree.

24/12/2018: An email from the DC was received, confirming the modification of "my" personal data.

The DC could not be persuaded to send the personal data to the adversary email. However, a request coming from the adversary email to modify the data was accepted, hence allowing an unauthorized change to the personal data of the subject. As we only exercised our right to modify personal information with *Fin_C*, it is inconclusive to know if more organizations are vulnerable to such attack in this scenario.

[7] Dates are using the little-endian notation

[8] We acknowledge that we did not expect to receive the data by postal mail as we were uncertain about the specific meaning of 'copy of the data'.

A.1.2 International financial institute: Fin_E

In the privacy statement of *Fin_E*, only an email address to send DRs to was provided, without information about the necessary credentials. Submission of the DR included the name and date of birth of the targeted individual and was carried out with the "Ringer" strategy.

20/11/2018: DR was sent by email.

21/11/2018: Email was received by the adversary, confirming the reception of the DR.

04/12/2018: A response from the DC on the original email was received, containing a summary of answers to the questions asked in the initial DR. In addition, they suggested us to "manually visit" the organization web pages in order to extract the necessary information. Article [2, Art. 68] states that data should be delivered "in a structured, commonly used, machine-readable and interoperable format" and hence, the response clearly violates this article.

17/12/2018: As a realistic adversary does not have the knowledge of the previous response received on the targeted individuals' email, we send them a persuasive reminder to indicate the approaching deadline according to [2, Art. 12-3].

17/12/2018: The DC responds to the adversary's email, stating that an answer to the DR was already offered and forwarded the email message from 04/12/2018 to us.

17/12/2018: Since the adversary is now aware of the original email being sent, we notify the DC of their violation of Article [2, Art. 68] and therefore, request the controller to send "my" personal data in a "machine-readable format".

18/12/2018: The targeted individual received a phone call on the number known by the DC. In this phone call, they verified the identity of the targeted individual by requesting the birthplace, original email address and specific account data.

27/12/2018: The targeted individual received the personal data (consisting of scanned documents) on the original email.

In this case, we argue that the DC did not receive any DR yet that explicitly mentioned the violation of [2, Art. 68], or the DC has no automatic process in place and attempts to eschew the DR by only providing limited information. Nonetheless in the end, additional verification methods were performed which are very difficult for an adversary to forge as it would require access to the phone number and specific knowledge related to the account of the targeted individual.

A.1.3 Logistics service: Trl_D

To submit a valid DR, the privacy policy of *Trl_D* states that a copy of the identity card, international passport or driving license is required. Additionally, the DC requested to censor sensitive information such as the photo and NRN. Similar to previous cases, this attack was performed with the "Ringer" strategy and all communication was done through the fake email address of the adversary.

19/11/2018: DR with the necessary credentials (as stated in the privacy policy) was submitted through a web form.

19/11/2018: Automatic reply was received, providing a ticket number.

22/11/2018: A response from the DC was received, asking if the email address of the targeted individual also had to be included into the data package. In other words, the DC indicates that there is an account with a different email address belonging to the targeted individual and therefore, requests if the personal information of this account should also be included.

22/11/2018: We replied that the email address is indeed an "old and unused one" and hence request personal information from that account.

18/12/2018: All personal data from the targeted individual, including additional data from other individuals with a seemingly similar name was received by the adversary in one large PDF file.

In terms of privacy, there are two breaches: (1) the impersonation strategy succeeded and (2) personal information of three additional users were leaked (link leakage). The first breach occurred rather quickly as the email of 22/11/2018 shows clear signs of the DC already assuming the identity of the DS without performing additional verification. The second breach indicates that including additional sensitive information from 3 other individuals is clearly also a privacy issue, albeit with a different impact compared to the previous cases. This type of breach would even exist if the DS would send a legitimate request, similarly to the publicly known 2018 Amazon Alexa data leak where a DS received voice recordings from an unrelated individual [19].

In our case, the occurrence of such mistake happened most likely due to an inaccurate query. With 2 of the 3 unrelated individuals, the cause was clear as the name of the unrelated individual was exactly the same as the name of the targeted individual. In the remaining case, the unrelated individual whose personal data was leaked had the following data:

- Name: A B
- Address: C-D

while the targeted individual had the following personal data:

- Name: B C
- Address: G

The “address” field of the unrelated individual was erroneously contained in the “name” field. Therefore, the resulting field of the unrelated individual became “A B C-D”, thus containing the string “B C”, which is precisely the name of the targeted individual.