In the Proceedings of the 32nd USENIX Security Symposium (USENIX Security'23). Anaheim, CA. August 2023.

Assessing Anonymity Techniques Employed in German Court Decisions: A De-Anonymization Experiment

Dominic Deuber Michael Keuchen
Friedrich-Alexander-Universität Erlangen-Nürnberg

Nicolas Christin
Carnegie Mellon University

Abstract

Democracy requires transparency. Consequently, courts of law must publish their decisions. At the same time, the interests of the persons involved in these court decisions must be protected. For this reason, court decisions in Europe are anonymized using a variety of techniques. To understand how well these techniques protect the persons involved, we conducted an empirical experiment with 54 law students, whom we asked to de-anonymize 50 German court decisions. We found that *all* anonymization techniques used in these court decisions were vulnerable, most notably the use of initials. Since even supposedly secure anonymization techniques proved vulnerable, our work empirically reveals the complexity involved in the anonymization of court decisions, and thus calls for further research to increase anonymity while preserving comprehensibility. Toward that end, we provide recommendations for improving anonymization quality. Finally, we provide an empirical notion of "reasonable effort," to flesh out the definition of anonymity in the legal context. In doing so, we bridge the gap between the technical and the legal understandings of anonymity.

1 Introduction

Crucial elements to a healthy democracy are accountability and transparency. While this naturally applies to lawmakers, it is also true for the judiciary, especially courts. This is why in jurisdictions within the European Union (EU), court decisions must be made accessible to the public [16]. In the US, or common law jurisdictions in general, the need for publication already arises from the paramount role of judicial precedent. In contrast to the US, the EU is more protective of the identities of people involved in the decision, mainly parties to the case, but also lawyers, expert witnesses and others [22, 31]. This leads to court decisions being anonymized prior to publication [15, 17]. In contrast to a technical definition, anonymity in the legal context is defined as not being de-anonymizable given reasonable effort in terms of time, costs and labor [39]. This definition brings up two questions. First, when is effort legally deemed reasonable [1]? Second, can information be de-anonymized given reasonable effort? This second question needs to be answered empirically by actually attempting de-anonymization. This is because general hypotheses about how easy de-anonymization could be are insufficient; *actual* effort, as opposed to *presumed* effort, is key. Technically de-anonymizable information can still be legally considered anonymous if the effort is deemed unreasonable. The results of a de-anonymization attempt might, in turn, be used to interpret the legal concept of "reasonable effort." This "reasonable effort" concept generalizes beyond German court decisions, as it is based on the General Data Protection Regulation (GDPR), applicable throughout the entire EU.

In the European practice of anonymizing court decisions, the most widely employed anonymization techniques rely on (i) using random or real initials, (ii) complete obfuscation, or (iii) replacement by role [40]. Opijnen et al. hypothesize that some of those techniques are vulnerable [40]. Trivial solutions, such as completely removing information that might lead to an identification of involved persons are undesirable, as they may impede the readability and comprehensibility of the decision, which need to be preserved [2]. The same holds for indiscriminate applications of differential privacy techniques as specific details in individual cases may be critical to the judgments rendered. For example, a court decision regarding a medical malpractice may hinge on whether a doctor (or an expert witness) is a specialist in a particular field.

In the absence of uniform regulations on what and how to anonymize, the practice is highly opaque, which may have tremendous ramifications for the privacy of the involved persons. Against this background, we identify the following research questions regarding the anonymization of court decisions.

RQ1: Which anonymized attributes or anonymization techniques are vulnerable?

RQ2: Which publicly available sources facilitate de-anonymization?

RQ3: Which insights into reasonable effort can de-anonymization attacks provide?

1.1 Our de-anonymization experiment

To answer these questions, we conducted a de-anonymization experiment analyzing court decisions from Germany. German court decisions are particularly suitable because, in general, in the absence of uniform regulations, each court decides on an individual basis what should be anonymized and how. As a consequence, German court decisions reflect all the anonymization techniques commonly utilized in European court decisions. Hence, our results are also relevant to anonymization of court decisions in Europe.

The experiment is structured as follows. We recruited a number of law students, presented them with selected court decisions, and asked them to de-anonymize them with the help of the internet. The decisions were selected to include a variety of anonymization techniques and anonymized attributes. To answer our research questions, we had participants document their findings. In addition, we recorded their internet activities.

Our resulting contributions are 1) to provide the first empirical analysis of different anonymization techniques in court decisions; 2) to assess the legal term *reasonable effort* by providing comprehensive quantitative and qualitative insights into the context of the anonymization of court decisions; and 3) to provide practical guidance on how to anonymize court decisions in a way that renders de-anonymization considerably more difficult.

1.2 Related Work

Anonymization of legal data Both the anonymization and publication of legal data are subjects of current research [2, 18, 26, 36, 40]. While these efforts suggest that the techniques currently used for anonymization are flawed [40] and try to justify this position with examples [2, 18], they do not provide the necessary empirical support.

Re-identification in medical data Re-identification attacks on or using medical data are the subject of intense research [7, 8, 19, 21, 29, 37, 49, 50]. The identification of a record in a structured medical dataset is arguably unverifiable, as the datasets are often incomplete and represent only a fraction of the population [6, 33, 47]. However, several researchers have characterized the uniqueness of attribute combinations in the population [37, 45, 48] and have shown that only few attributes are enough to uniquely identify an individual.

In contrast to this line of research, we study legal data, which differs in many respects from medical data as discussed by Csányi et al. [18]. In particular, legal decisions might involve or describe rarely occurring events that could appear in the news. Also, compared to typical medical datasets, legal data can contain far more attributes, and as such, are high-dimensional. Because of the rare events and relations between involved parties, legal data are also sparse [18]. In general, high-dimensional sparse data have been shown to be vulnerable to de-anonymization attacks [20, 35]. Prominent instances of such attacks include those on the Netflix Prize dataset and the AOL search query logs [5, 34].

Assessing effort Some related work provides insights into de-anonymization efforts [8, 52], including court decisions [56]. Tudor, Cornish, and Spicer de-anonymized UK census data [52] using a so-called “motivated intruder” test [27], a practical method for studying de-anonymization risks. In a motivated intruder test, a reasonably competent person tries to identify an individual from anonymized data only using publicly available sources, and without prior or specialist knowledge [27]. Although census data, like medical data, cannot be compared with legal decisions, this work allows us to put our analysis of reasonable effort in context. Branson et al. conducted an intruder test to evaluate anonymization guidelines issued by the European Medicines Agency [8]. Even though this work evaluates anonymization guidelines, it only focuses on the general re-identification risk and not on specific techniques utilized to anonymize text.

Perhaps most closely related to this paper, Vokinger and Mühlematter [56] performed a re-identification attack on selected decisions of the Federal Supreme Court of Switzerland. In contrast to our experiment, the proceedings were limited to those initiated by pharmaceutical companies against price-fixing orders issued by the Federal Office of Public Health. Furthermore, they re-identified anonymized drugs and the complaining pharmaceutical companies by linking non-anonymized information in the decision with official data from the Federal Office of Public Health. As a consequence, Vokinger and Mühlematter did not study different anonymization techniques employed in the legal context and only focused on one specific source during de-anonymization.

2 Legal Background

We next explain why there is a need for publication and anonymization of court decisions, and provide an overview of the anonymization practice in Germany.

Obligation to publish German case law developed an obligation to publish court decisions based on the constitutional principle of democracy and the rule of law [14, 15]. Court decisions must be published to inform the public, media, legal professions and other courts about the developments of the interpretation of legislation. The argument is that publicly available, comprehensible court decisions are necessary for citizens to adapt their legal behavior [10]. In addition, transparent jurisprudence is a prerequisite for control by the people, public discourse and responses by the legislator [15]. In the last decade, only 2.3% of German court decisions were published on official and freely accessible portals [30], showing there is still huge potential for publication.

Obligation to anonymize Nevertheless, the legitimate public interest in information is limited by the legally protected interests of natural and legal persons involved in a court decision [22]. These legally protected interests include the right to informational self-determination [9], the right to company and business secrecy, to tax secrecy, to data protection, and to protection of a company's reputation [1]. This is why court decisions need to be anonymized prior to publication [15]. In Germany, there is no *single* statute that conclusively defines which features need to be anonymized or which anonymization techniques should be utilized [40].

Legal practice of anonymization The lack of a uniform regulation on what and how to anonymize leads to inconsistencies in legal practice. As a result, some courts created their own internal anonymization guidelines [40]. However, the absence of guidelines in other courts results in completely different approaches. Often, anonymization is done manually, takes a very long time and is highly prone to errors [26]. Semi-automatic approaches based on search and replacement are less frequent. Approaches to anonymization are inconsistent to the point that several different techniques may be used in a single court decision. In addition, the extent and degree of anonymization also vary from court to court, which is why the anonymized texts retain different amounts of information.

Inconsistencies in the legal practice aside, anonymization of court decisions is non-trivial [18]. Indeed, from an information-theoretical standpoint, completely omitting the information would be ideal. However, this approach is inadequate, because the purpose of publishing court decisions is to inform. Thus, the legal argumentation has to be preserved [2]. This requirement causes a conflict between readability and comprehensibility on the one hand and effective protection of the legitimate interests of the involved persons on the other hand [26, 36]. In fact, for a given case, a judge may even consider the comprehensibility of the court decision to be more important than its full anonymization [38].

In the future, anonymization will arguably have to be automated to handle the large number of decisions that are currently not published. Also, an automated anonymization must preserve readability and comprehensibility, protect the involved persons, while simultaneously allowing a judge to perform individual adjustments.

The concept of reasonable effort Information under German and EU law is deemed anonymized if de-anonymization can only be achieved with "*unreasonable effort*." More precisely, under German law, court decisions are considered anonymous if individual information about personal and factual circumstances of the anonymized entities cannot be determined at all or *only* with unreasonably large effort in terms of *time, costs and labor* [39]. At the EU level, GDPR Recital 26 defines the identifiability of a person very similarly.

Unfortunately, when exactly effort is to be considered reasonable is not clear. Case law only provides a couple of insights. First, the greater the potential value of the anonymized information, the more effort is considered to be reasonable [13]. Second, for the identification of a person, additional knowledge from generally accessible sources can be used as long as the effort remains reasonable [39, 55]. Furthermore, using legal means to acquire additional information from third parties [11, 24], e.g., internet service providers, is admissible and reasonable. Beyond these case law insights, what constitutes reasonable effort in terms of time, cost and labor is not precisely specified. Moreover, even assuming a reasonable effort, one must still empirically assess whether de-anonymization is possible.

De-anonymization risks The increasing online availability of court decisions, and more generally, the amount of data available on the internet, creates new opportunities to study court decisions more extensively [3]. Singling out individuals in cases with high media coverage, criminal convictions, or the involvement of prominent persons, becomes potentially easier. Likewise, large-scale statistical analyses of certain procedural roles become possible, including, for example, how often an insurance company or a brand is being sued. These derived analyses can in turn be used to influence the public image of individuals or companies. For instance, rating services can include these analyses in their portfolios to estimate how often a particular insurance company needs to be sued before it will pay. These analyses can also lead to unlawful profiling, discrimination and multiple infringements on the rights of legal and natural persons [22]. As discussed earlier, de-anonymization risks stemming from publicly available court decisions differ from other (e.g., medical [57]) data sources, as court decisions contain a large amount of information, such as descriptive statements concerning the facts of a case [22]. This information, combined with each other or with external sources, can lead to de-anonymization [2, 4].

3 Terminology and Attack Model

We next present our terminology which adapts the general anonymity terminology of Pfitzmann and Hansen [42] to the legal context.

Every *court decision* defines a set of *involved entities* that can either be *natural* or *legal persons*. "Involved" means that the entity was either directly involved in the facts of the case or in its proceedings. That is, in addition to the claimant and respondent, involved entities include lawyers, witnesses, experts, etc. The protection of these entities' interests must be ensured [22] and thus, the court decision must be anonymized prior to publication [15]. Every entity is defined by numerous (combined) attributes. For example, natural persons are defined by their name, date of birth, profession, title or address. We refine the terminology of Pfitzmann and Hansen [42] by distinguishing *attributes*, *sub-attributes* and *attribute values*. An attribute defines a set of (attribute) values, e.g. the set of all first names. A sub-attribute defines a subset of an attribute set, e.g. all first names that start with the letter "E." Combined (sub-)attributes are the Cartesian product of two or more (sub-)attributes, e.g., names of natural persons are the combined attribute based on first, last, and possibly middle names.

Attribute values appear in the original decision as *attribute-value strings*, potentially comprised of sub-strings separated by white spaces. These attribute values are usually what is anonymized in a decision.

Definition 1 (Anonymized string) *Any unique appearance of anonymized information in a decision is an* anonymized string.

Anonymized strings can also be comprised of sub-strings separated by white spaces. We define anonymity at the attribute level. An attribute value is deemed anonymous if it is not identifiable within a larger attribute set, i.e., if it is not *uniquely* characterized within that set [42]. *De-anonymization* refers to the reduction of the anonymity set size.

3.1 Attack model

The attack model stems from our research questions. To detect which anonymized attributes or anonymization techniques are vulnerable (RQ1), our model assumes an attacker that aims to de-anonymize as many anonymized strings as possible in a given court decision. In other words, the attacker's goal is to retrieve the original attribute-value strings, before they were anonymized. Vulnerable attributes or techniques can only be considered de-anonymized *in the legal sense* if the effort of the attacker was reasonable (RQ3). Thus, we greatly restrict our attacker in terms of time, costs and labor. Our attack model assumes a single human attacker with access limited to publicly available sources (RQ2) on the internet. The attacker does not have any budget to access non-public sources, e.g., paywalled information. The attacker has complete freedom in determining which anonymized strings they try to de-anonymize and how, but only has a fixed, finite amount of time to carry out the attack. Finally, the attacker only receives minimum wage for their time spent on de-anonymization. Our minimalist approach is intended to rule out situations in which time, labor or costs are unreasonable.

3.2 Attributes

Throughout the experiment, we considered the following attributes in line with existing work on the anonymization of German court decisions [1, 26, 36].

Names The first and most important attribute is the name of an involved entity. We further distinguish names of natural and legal persons. At a minimum, names must be anonymized in a decision, as they usually directly identify the entity [4].

Locations Location is also often anonymized to make entity de-anonymization harder. For instance, not knowing where a case took place makes it harder to identify a local expert witness. We distinguish street names, cities and countries.

URLs We include URLs, as a de-anonymized URL could be a good starting point for further de-anonymization.

Authorities Authorities refer to official bodies, such as courts, administrations, youth welfare offices, district offices and so on. Authorities could be listed under legal entities (and thus under names), but we argue that we need this distinction, because one can almost always find online information on an authority, and their names frequently include locations.

Dates, IDs, Miscellaneous We also considered dates and ID numbers (such as account numbers, license plates, etc.). We denoted as "miscellaneous" anything that does not fall into the previously listed attributes and occurred rarely; for example, contact data, product names or university degrees.

3.3 Anonymization techniques

We distinguish four different categories of anonymization techniques. We adapt the techniques proposed by Mamede, Baptista, and Dias [32] as we are faced with anonymized decisions and not with the anonymization task itself. The most prevalent techniques to anonymize court decisions in Europe are (i) random or real initials, (ii) complete obfuscation, or (iii) replacement by role [40]. Our categories allow us to represent these techniques as shown below.

Suppression *Suppression* refers to the replacement of the attribute-value string, one up to all sub-strings or individual characters, with (a) neutral expression(s). An expression is neutral if it has been chosen independently from the term it replaces. For example, "Los Angeles" could be replaced with "...," "... ....," "XXX," "P. Q.," or "A". The date "07.06.2021" could be changed to "00.00.2021".

Omission *Omission* refers to the (partial) omission of the attribute-value string, sub-strings or individual characters. For example, "John Doe" could become "J. D." or just "J." "Abbey Street" could result in "A. Street".

A priori omission *A priori omission* is a special case of omission, which we list separately because it plays a crucial role in our experiment and in legal practice. Here, the attribute-value string, e.g., name of the claimant or respondent, is omitted *during writing* of the decision. The remaining string only denotes an attribute class, e.g., claimant or respondent, and possibly a unique identifier for distinction. In legal practice, this technique is only used for specific parties of the case.

Tagging *Tagging* refers to the replacement of the entire attribute-value string, or one up to all sub-strings with a string that explicitly denotes an attribute class and might contain a unique identifier for distinction. For example, "Los Angeles" could be replaced by "City1" or "Apple Inc." could be replaced with "Company1 Inc." The main difference with *a priori* omission is that the attribute-value string is not omitted during writing but replaced *a posteriori*.

Finally, *random substitution* refers to the replacement of the attribute-value string with another attribute-value string randomly chosen from the same attribute set. We do not consider this technique any further, because 1) random substitution only plays a marginal role in the anonymization of European court decisions [40], and 2) one cannot reliably determine when random substitution is used, as it is indistinguishable from the original attribute-value string.

The most common anonymization techniques in European court decisions map to our categorization as follows: random initials correspond to suppression, real initials to omission, replacement by role to tagging, and complete obfuscation to either suppression or omission.

3.3.1 Partially preserving vs. non-preserving techniques

We further distinguish anonymization techniques by how much information of the original attribute-value string they preserve. Any anonymization technique can be either *partially preserving* or *non-preserving*. Completely omitting the attribute-value string (e.g., as in a priori omission) is non-preserving, while any other form of omission always preserves some information about the attribute-value string, namely the information not omitted. Suppression can be either non- or partially preserving. If the entire attribute-value string is replaced with a neutral expression, suppression is non-preserving. In all other cases, suppression preserves some information about the structure of the attribute-value string. For example, replacing "New York City" with "... ...." preserves the number of sub-strings and their length, even though each individual character has been replaced with a neutral expression. If tagging only replaces some sub-string (e.g., "Company1, Inc."), then it is partially preserving; otherwise, it is non-preserving.

4 Experiment Design

We designed an experiment that implements our attack model (see Section 3.1). After providing a general overview, we elaborate on the individual design choices. We selected 50 anonymized court decisions and recruited 54 participants as attackers. Each participant worked for a maximum of 3 hours and had (at most) 35 minutes for each attack. Accounting for the initial introduction and instructions, we aimed for each participant to attack at least four decisions. The participants had to document their results in an application while we also recorded their browser activities.

4.1 Filtering court decisions

We used the following process to select the German-language court decisions presented to the participants. First, we downloaded all publicly available court decisions of the 16 German states up to March 2021 from their respective portals, obtaining a total of approximately 356 000 decisions.[1] The details of this process are described in another work [30].

Year and jurisdiction We then filtered the decisions by year and jurisdiction. We only considered decisions from 2016 to 2020 because certain German states did not publish anything older. We also restricted decisions to those from "ordinary courts" that deal with civil and criminal cases, and thus excluded so-called special jurisdictions that deal with labor, financial, social and administrative cases, for two reasons. First, not all German states publish decisions from special jurisdictions on their portals. Second, we intended to prevent undesirable effects that the special jurisdictions might introduce, notably the fact that the attributes of the involved entities in these decisions might largely differ from those in decisions of ordinary jurisdiction. All in all, this filtering stage produced approximately 35 000 decisions.

Length To ensure that our participants had enough time to read *and* attempt to de-anonymize the decision, we did the following. We tokenized the text of the court decisions, ignoring punctuation, to estimate their word counts, and chose to only keep decisions between 2 000 and 4 000 tokens. The average adult reading rate for German text is 179 (± 17) words per minute (wpm) [51]. However, court decisions usually require careful and thus slower reading [44] and often use elaborate language. Recall that we only give participants a total of 35 minutes per decision (see Section 4.3). With 2 000 tokens and 179 wpm, one could presumably read the entire decision in around 11 minutes; with 4 000 tokens at least half of it. We made explicit in the experiment description we provided to our participants that the focus is not on the legal analysis of the decision but its de-anonymization. After this filtering stage, we had about 10 000 decisions left.

Anonymization techniques and attributes We then used regular expressions to preselect the 50 decisions that we presented to the participants. Our regular expressions consisted of expressions capturing (sub-) attribute(s) string(s) combined

[1] We did not consider decisions from the federal courts, as these often receive a great deal of media attention, which might have distorted the results.

Table 1: Technique distribution, broken down between partially and non-preserving techniques (see Section 3.3.1)

Technique	**Occurrence** (Count)	(%)
A priori omission, non-preserving	110	22.7
Omission, partially preserving	107	22.1
Suppression, non-preserving	142	29.3
Suppression, partially preserving	60	12.4
Tagging, non-preserving	45	9.3
Tagging, partially preserving	20	4.1
Total	*484*	*100*

Table 2: Attribute distribution

Attribute	**Occurrence** (Count)	(%)
Name (Natural person)	178	36.8
Name (Legal person)	99	20.5
Location (Street)	24	5.0
Location (City)	61	12.6
Location (Country)	9	1.9
Authority	27	5.6
Date	12	2.5
URL	31	6.4
ID	22	4.5
Miscellaneous	21	4.3
Total	*484*	*100*

with expressions capturing the techniques. We included all techniques as described in Section 3.3. For attributes, we focused on names, locations and URLs as described in Section 3.2. Finally, we randomly drew 50 decisions, and replaced decisions with few or overrepresented anonymization techniques and attributes until our selection included all techniques and attributes.

Classifying techniques and attributes As regular expressions cannot detect all anonymized strings, we only used them to preselect the decisions. We obtained the final distribution of attributes and techniques by manually inspecting each decision. As we did not have access to the original decision before anonymization, we did not know which technique(s) were employed and had to carefully reverse-engineer them. For example, "X Street" might be Xavier Street, or X might be a neutral expression. We follow a conservative approach, whereby we take into account the entire decision and other decisions from the same court or jurisdiction. We interpret initials (omission) as random (suppression) if there is no evidence to the contrary.

Tables 1 and 2 show the distributions of techniques and attributes, respectively. 484 anonymized strings were distributed among the 50 decisions, for an average of 9.7 per decision. The *tagging* technique occurs relatively rarely, while on the attribute side, names are frequently present. The distribution of techniques mirrors that reported by Opijnen et al. [40] for European decisions, although we ensured – through our selection process – that tagging techniques were used for more than 10 percent of all anonymized strings.

Limitations A limitation of our approach is that the techniques and attributes cannot be examined in isolation and independently of non-anonymized attributes. However, we ruled out other design choices. Creating artificial decisions with an equal distribution of techniques and attributes would have been ineffective, as de-anonymization of such decisions would have been impossible due to the lack of publicly available sources. Likewise, creating a dedicated version of an existing decision for each technique would have been impractical. First, we had no access to the original decisions and thus would have missed the information necessary to correctly model *omission, partially preserving*. Second, we would have been unable to prevent the participants from discovering the existing decision on the internet, thereby tainting our results. In fact, doing so would have been very easy, using the case number, its wording or the facts of the case.

4.2 Participants

We recruited 54 law students from a German university (hereinafter: the University) to participate in this experiment. We deliberately decided to limit ourselves to law students, based on the outcome of pilot testing, for two reasons. First, law students have experience in reading decisions due to their studies. Second, there were significant differences in the way students from other disciplines read and understood decisions, which could have affected results.

Participation was voluntary and remunerated with a EUR 30 Amazon voucher. Participants ranged in age from 18 to 32, with a mean of 22.5. Nearly two-thirds of the participants were female (35), the others male (19). This largely corresponds to the gender distribution of law students and the age distribution of all students at the University. On average, participants were in their sixth semester (with a minimum of 2 and a maximum of 14).[2] All but one participant reported experience with the Windows operating system and more than 85% with Google Chrome. We can thus neglect the effect of operating system or browser difficulties may have had on our results. In addition, we used a five-point Likert scale (1: not at all, 5: very much) to ask our participants if they had experience with 1) reading court decisisons and 2)

[2] As the qualification phase for admission to all legal professions in Germany generally takes seven years, the participants were still at an early stage.

data (de-)anonymization. Participants generally did not have much experience with (de-)anonymization (average: 1.6), but did have significant experience with reading court decisions (average: 3.4).

Finally, de-anonymization skills might greatly differ from one participant to the next. To mitigate this, we had several participants attack the same decision. As the experiment took place on several days, the participants signed an agreement not to disclose our experiment design.

4.3 Time limits

The experiment ran for at most three hours, including initial introductions and instructions. Participants were given only 35 minutes for each attack, so that they could carry out multiple attacks (ideally at least four), during the three-hour period. After 35 minutes, the participants were informed that time was up and were not allowed to continue. However, they were asked to document any findings, even preliminary or incomplete. Participants could terminate the de-anonymization of the current decision at any time in the application. If they elected to do so, we required them to justify their decision. Such premature terminations could increase the total number of attacks a given participant was engaging in, as long as they stayed within the three-hour time limit.

Limitations We restricted the total time to three hours to limit participant fatigue. As each participant performed several attacks, we assumed a potential a learning effect, i.e., a participant might figure out how to effectively de-anonymize and apply this knowledge to subsequent attacks. On the other hand, this acquired knowledge might not be transferable to another attack. An example would be that a participant considers all characters of an anonymized string to be initials, which only holds in the case of omissions, but not with suppressions. We mitigated potential learning effects with the results by randomizing treatments. Specifically, as a decision was attacked by several participants, we ensured that for one participant it was their first attacked decision, for another their second, and so on. In case multiple decisions were available for a participant's *n*-th attack, these were chosen randomly. Finally, because of the time limit, some de-anonymizations might only be partial. That said, because multiple attackers worked independently on the same decisions and chose which strings to attack, we expected a large number of anonymized strings to be attacked.

4.4 Experimental setup

The experiment was conducted at the University, where each participant worked alone in a room on a university-provided computer with two screens. The participants first received instructions (Section 4.4), then were given access to an application guiding them through the experiment (Section 4.4) and finally performed their attacks in Chrome. They used the application to document their results (Section 4.4).

Instructions The experiment started with a verbal introduction and necessary explanations about the data collected and data protection measures. The application then collected basic demographic information and experience as described in Section 4.2, gave a detailed explanation of the experimental procedure, and instructions on how to document any results. Participants could subsequently review these instructions at any time. We emphasized, both in the written instructions and in our verbal briefing, that we were not only interested in successful de-anonymizations but also in unsuccessful attempts. We purposely did not stipulate any requirements or provide any guidance on how to proceed with de-anonymization. To avoid biasing the participants in their approach and their selection of anonymized strings to attack, we then simply showed them the anonymized decision. Participants could decide what they wanted to de-anonymize, in what order, and whether they wanted to record the anonymized strings in the application immediately or only after they had found something.

Experimental application We designed a Windows application specifically for this experiment. Participants could call a researcher from the application to deal with technical difficulties and could also pause the experiment for bathroom breaks. The application recorded the break times to take them into account accordingly in the evaluation of reasonable effort (see Section 6.5). However, the main purpose of the application was to document the results.

As we had no access to the original decision, we needed another way to assess whether a de-anonymization was successful. To do so, we first required the participants to extensively document, in the application, their findings by providing the de-anonymized information, a confidence score, links to their sources and a justification. We used the confidence score to limit ourselves to de-anonymizations the participants were certain about. Then, we tried to verify the participants' results, using the provided links and justification as a starting point (Section 6.2).

Participants documented their results in a separate table for each attack. Participants created the table rows, which could be edited or deleted during the attack. However, the columns were fixed as follows. The documentation interface is shown in the appendix (Figure 5).

Anonymized string. In this column, the participants needed to enter the anonymized string as it appeared in the decision.

De-anonymized information. If de-anonymization was successful, participants entered the de-anonymized string here. Otherwise, this column could be left blank.

Confidence. This column allowed indicating the degree of confidence on a six-step scale (1: very uncertain; 6: very certain). We purposely used an even scale to avoid a "neutral" midpoint. This confidence level had different meanings depending on whether the de-anonymized information column was filled. If de-anonymized information was entered, the degree of confidence denoted how certain the participants were about their results. If nothing was entered, the degree of confidence denoted how certain the participants were that de-anonymization of that string was not possible. Participants had to either fill in the degree of confidence or delete the entire row at the end of the time alloted to the corresponding decision. Only in this way could we distinguish whether de-anonymization was attempted at all, or whether the table was merely used to document anonymized strings.

Links. If participants filled out the *De-anonymized information* column, they had to enter in this column links to one or more sources they derived their de-anonymization from.

Justification. Here participants had to justify why they either thought they achieved de-anonymization or why they could not achieve it. In contrast to the degree of confidence, the justification provides a qualitative insight.

Customized Google Chrome Participants had access to the entire internet accessible via the University network, including online library access. In addition, the participants were informed that no login data would be stored, so they could log into their social media accounts without any concerns. The participants were only allowed to access the internet via a customized version of Google Chrome that recorded participants' browser activity using the Security Behavior Observatory (SBO [23]) browser sensor [41].

5 Ethical Considerations

In the course of the experiment, we not only collected personal data of the participants, but also of the natural and legal persons involved in the attacked decisions. Before the start of the experiment, all participants consented to the processing of their data both in writing and again in the application. To protect the participants and the persons involved in the attacked decisions, we implemented the following safeguards. All data processing within the scope of the experiment complied with the GDPR and had been approved by the University's Data Protection Officer. In particular, we encrypted any personal data and restricted access to the research team and a minimal support staff. Any staff who had access was bound to keep all personal data they acquired in the course of their work confidential, by the terms of their employment contract. Likewise, the participants signed non-disclosure agreements to protect the natural and legal persons involved in the attacked decisions. We could not inform these involved entities, as we had either no or very limited contact information for them. We thus cannot publish any information that would divulge the 50 decisions attacked in the experiment. In short, we attempted to balance our research exposition with the protection of the involved persons. Finally, one of the co-authors, at a separate university, never had access to individual data. The relevant IRB confirmed in writing no application was necessary.

6 Findings

Before we discuss how our experimental results informed answers to our research questions, we first clarify how we categorized the publicly available sources (Section 6.1) and how we verified the results of the participants (Section 6.2). In the (general) absence of a ground truth, these preliminaries are important to ensure the robustness of our findings.

6.1 Source categorization

Based on the outcome of our experiment, we categorized the publicly available sources that the participants reported using as follows.

Attacked decision Sometimes, the attacked decision *itself* leaks information. For example, the attribute-value string may be accidentally anonymized in all but one place.

News This category includes any news articles from media and journalists as well as blog posts from any other page reporting new incidents or developments. For example, a lawyer's blog page could discuss a recent decision.

Personal page This category denotes information that was published on the webpage of legal or natural persons – i.e., individuals or organizations – involved in the decision.

Search engine This category includes all web search engines (Google, Bing, DuckDuckGo, etc.).

Mapping This category represents searchable mapping services, such as Google Maps, as well as depictions of a map.

Wiki This category covers general pages that provide information and gather knowledge – examples include statistical pages, wikis, journals or street directories. However, these directories are not run by any official institution.

Business directory In contrast to *news* and *wiki*, this category is independent of recent events. It includes reviews or listings of companies, professions (e.g., experts, doctors, lawyers), services, products or people. Thus, the category includes business directories, expert portals, business search engines (e.g., Yelp) or yellow pages.

Official directory This category covers public sources provided by official institutions, such as courts or administrations (e.g., the European Patent Office). These sources are highly reliable and complete.

Other decision This category contains previous or subsequent court decisions related to the attacked decision.

Miscellaneous This includes all remaining, rarely used types of sources that do not fit into any of the other categories – e.g., user profiles in a forum or marketplaces.

6.2 Result verification

Since we did not have access to a "ground truth," but merely to the participants' findings, justifications, and sources, we needed to verify whether they actually achieved de-anonymization. Our starting point is the *potential* de-anonymizations participants reported.

Definition 2 (Potential de-anonymization) *A* potential de-anonymization *is a single participant's reported de-anonymization concerning an anonymized string (Definition 1) with a degree of confidence greater than three.*

We only considered de-anonymizations with a degree of confidence greater than three, as we required the participants to be certain about their findings. As we had no access to the original decision, we too needed to rely on publicly available sources. However, unlike the participants, we were not limited to 35 minutes and have legal and technical knowledge that exceeds that of the average law student. To verify the participant's potential de-anonymizations, two of the authors searched for sources and justifications for every anonymized string with at least one potential de-anonymization. The sources, as described above, should reduce the anonymity set size to one. The justifications argue why the de-anonymization worked and ruled out false positives. We categorized justifications as arguments of the following types.

Legal This refers to any information that follows directly from the law. Specifically, the law dictates which court has jurisdiction in which locations, which judges hear which case, and the organizational structure of companies. Consider a decision that mentions "City B" (anonymized by omission). If only one of the cities where the deciding court has jurisdiction starts with a "B," we can deduce which city "B" refers to.

Self This refers to any information that follows directly from the attacked decision, such as our earlier example where the information was left unredacted in one (or more) occurrence(s).

Ground truth Ground truth refers to the original decision before anonymization. Ground truth is generally unavailable, but, in one case, one of the parties to the case published the original decision on the internet.

Geographic exclusivity This category refers to the limited anonymity set sizes introduced by geographic exclusivity. For example, a country has only one capital city. Cities might have only one street whose name begins with a rare letter combination, e.g., "R.-L.-Platz."[3]

Unique inference This refers to other non-anonymized information in the decision that is so specific – by itself or in combination with information from other sources – that the anonymity set directly reduces to a singleton. For example, an (anonymous) athlete ranked fifth at a very specific tournament in a given year, or the author of a book whose title is made explicit, can be uniquely identified.

Table 3: Sources and justifications used in verification of potential de-anonymizations. (Key: L: legal; S: self; GT: ground truth; GE: geographic exclusivity; UI: unique inference.)

	L	S	GT	GE	UI	*Total*
Attacked decision	6	15	0	2	0	*23*
News	0	0	0	0	13	*13*
Personal page	1	0	0	0	19	*20*
Search engine	0	0	0	1	2	*3*
Mapping	0	0	0	9	5	*14*
Wiki	0	0	0	1	7	*8*
Business directory	1	0	0	0	3	*4*
Official directory	16	0	0	0	6	*27*
Other decision	4	0	10	1	10	*25*
Miscellaneous	0	0	0	0	6	*6*
Total	*28*	*15*	*10*	*14*	*71*	

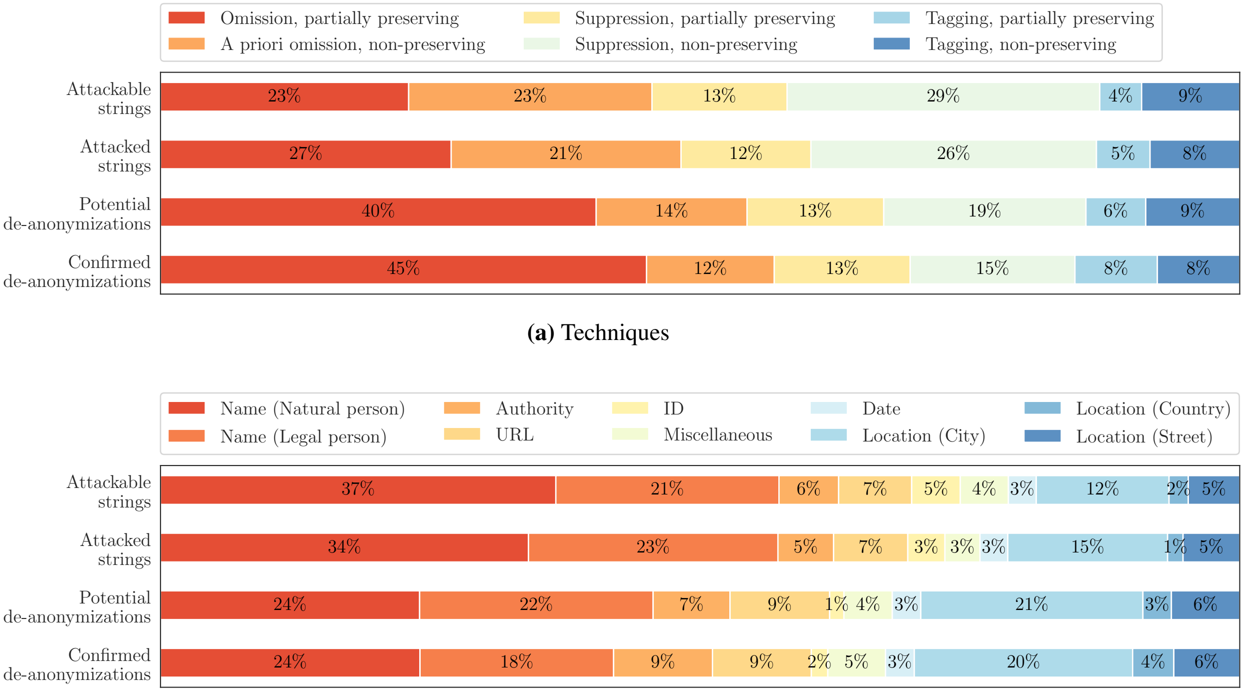
Verification results. Our participants potentially de-anonymized 184 (38%) of the 484 anonymized strings across the 50 decisions we presented to them. In 120 (65%) of these 184 strings, we could find one or more sources with appropriate justifications (and thus determine whether the de-anonymization succeeded).

Definition 3 (Confirmed de-anonymization) *A* confirmed de-anonymization *is a potential de-anonymization (Definition 2) that was verified and confirmed.*

Table 3 shows the sources we found along with their respective justifcations. More than half (51%) of the justifications come from *unique inference* arguments, while 20% resort to *legal* arguments. The other types of justifications (self, ground truth, geographic exclusivity) represent between 7 and 11% of all arguments. Taken together, the attacked decision itself, personal pages, official directories and other decisions represent almost two-thirds of all our sources, each ranging from 14 to 18%. News and mapping are around 10% each, while other sources rank between 2 and 6%.

Interestingly, in 18 of these 120 cases, we had more than one source and justification, as only a combination of sources and justifications allowed the verification. For example, a *legal* justification reduced the anonymity set size to two cities beginning with a "B"; and a *unique inference* argument further reduced the anonymity set to a singleton: the decision specified the city's population, which highly differed from that of the other candidate. Such cases are represented by multiple distinct entries in the table, which explains why the totals do not add up to the same numbers.

[3] "Platz" refers to town squares and is also the official address for the adjoining buildings in Germany.



(a) Techniques

(b) Attributes

Figure 1: Distributions of attackable strings, attacked strings, potential and confirmed de-anonymizations

6.3 De-anonymization results (RQ1)

Having been able to verify our participants' potential de-anonymizations, we next turn to the results of their attacks. Our 54 participants attacked on average 4.3 decisions each. Thus, in total, they carried out 231 decision-level attacks corresponding to 2 244 possible attempts to de-anonymize strings.[4] We will refer to these as "*attackable strings*."

Out of these 2 244 possible attempts, our participants actually attempted 1 047 de-anonymizations. We will refer to these as "*attacked strings*."

379 of these attacked strings resulted in *potential de-anonymizations* (Definition 2), of which almost 95% potentially reduced the anonymity set to a singleton, i.e. reported exactly one attribute-value string. For these 379 potential de-anonymizations, our participants reported an average degree of confidence of 5.4 (i.e., high). We could verify 294 of these 379 potential de-anonymizations, and confirmed 262 of them.

Figures 1a and 1b show the technique and attribute distributions of attackable and attacked strings as well as potential and confirmed de-anonymizations. The total numbers can be found in the appendix (Table 5). There, we also provide a histogram showing the number of attacked strings and confirmed de-anonymizations per decision (Figure 6).

Figure 1a indicates that all anonymization techniques were subject to de-anonymizations. We observe statistically significant differences between attackable strings and confirmed de-anonymizations ($\chi^2_5 = 81.9$, $p < 10^{-15}$). In particular, the percentage of partially-preserving omission and tagging techniques almost doubled – meaning those techniques are disproportionately successfully attacked. Non-preserving suppression and a priori omission, on the other hand, were cut in half. Last, partially-preserving suppression, and non-preserving tagging remained nearly constant.

Figure 1b also shows that confirmed de-anonymizations apply to all anonymized attributes. We again observe statistically significant differences between attackable strings and confirmed de-anonymizations ($\chi^2_9 = 40.9$, $p < 10^{-5}$). Most attributes remained roughly at the same proportion across both sets, with the exception of natural persons and IDs, which were markedly less present in confirmed de-anonymizations, and cities and countries, which were far more represented.

6.4 Reported publicly available sources (RQ2)

Figure 2 shows what publicly available sources the participants reported in the case of potential de-anonymizations. News sites and personal pages accounted for the largest share, at 23% and 21%, respectively. Other published court decisions, business directories and wikis each accounted for 9%. Official

[4] Each of the 484 anonymized strings (Definition 1) in the 50 decisions could potentially be subject to multiple attacks.

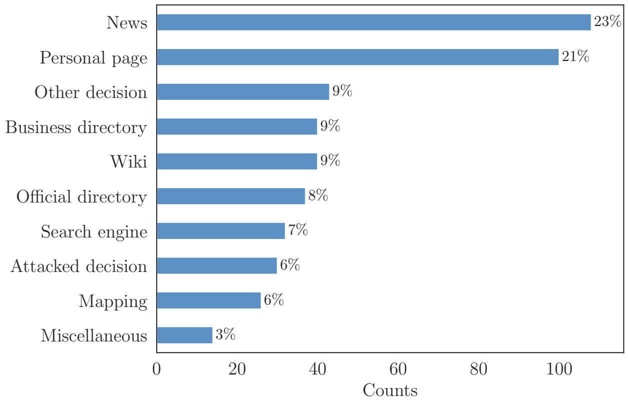


Figure 2: Publicly available sources participants reported in potential de-anonymizations

directories, search engines, the attacked decision and mappings ranged between 6% and 8%. 470 sources were reported in total. This exceeded the 379 potential de-anonymizations, as in 70 cases participants reported more than one source.

6.5 Measured effort (RQ3)

Time The average time expenditure per attack was 32.7 minutes. Attack time is divided approximately equally between using the application, examining the decision, Google, and other sources, with slightly more time spent on the latter (28%). This breakdown includes premature terminations, as these were explicitly foreseen in the experiment design. The exact timings with respect to premature terminations can be found in the appendix (Table 4). As the participants had two screens to work with, the times spent are an upper bound on actual participant attention time.

Figure 3 shows the times at which potential de-anonymizations were reported after the start of an attack. The curved line is a kernel density plot. No potential de-anonymizations were reported in the first two minutes after the start. Reported de-anonymizations increased and reached a peak at $t = 15$ minutes, before dropping slightly until $t = 24$ minutes. After that, the reported potential de-anonymizations increased slightly towards the end of the attacks, with peaks at $t = 30$ and $t = 36$ minutes. Potential de-anonymizations were still documented 43 minutes after the start, as participants were still allowed to document results after the maximum attack time of 35 minutes had elapsed. The peak at $t = 36$ is likely due to people documenting everything immediately after the conclusion of the experiment, represented by the red line in Figure 3.

Labor To assess labor, we focus on the average attack. The average attack contained 9.7 attackable strings, of which 4.5 were attacked; this yielded 1.6 potential de-anonymizations, of which 1.1 were confirmed. Each of the 484 anonymized strings was attacked by 4.6 participants on average. Figure 4 shows for each decision how many participants attacked the decision and how many of them achieved confirmed de-anonymizations. For readability, decision numbers are ordered from the least to the most vulnerable.

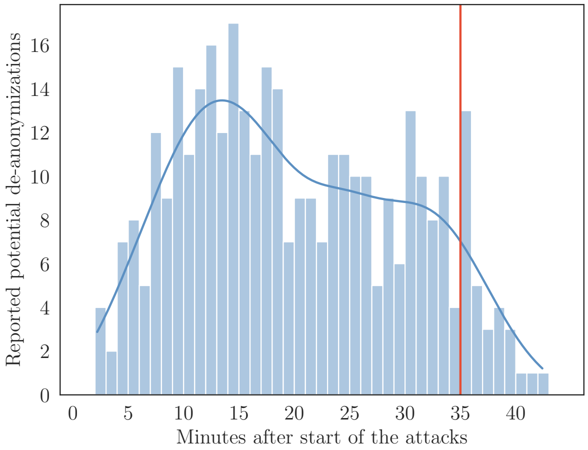


Figure 3: Potential de-anonymizations participants reported in all attacks

As we had assumed, participants' performance varied. We first computed the minimal number of participants needed to achieve the same level of performance that we overall observed in any given decision. We considered all possible combinations of participants that attacked a decision and determined the smallest combination that contained all confirmed de-anonymizations possible for that decision. We found that the smallest combination consisted of 1.3 participants on average, while for no decision would more than two participants have been needed to achieve all confirmed de-anonymizations.

These numbers assumed an optimal participant combination. We also computed how random combinations would perform. For each decision, a single, two, three and four random participants would respectively achieve on average 45%, 70%, 85% and 95% of all possible confirmed de-anonymizations.

Monetary costs Each of the 54 participants received a EUR 30 Amazon voucher. The entire experiment thus cost EUR 1 620. With 231 attacks on entire decisions, the average cost per attacked decision is around EUR 7. To achieve 95% confirmed de-anonymizations, we would need four independent attacks, which would bring the cost to EUR 28 for a given decision. This figure is a lower bound, as it does not include costs for recruiting and instructing the participants, securing a physical location, electricity costs, etc.

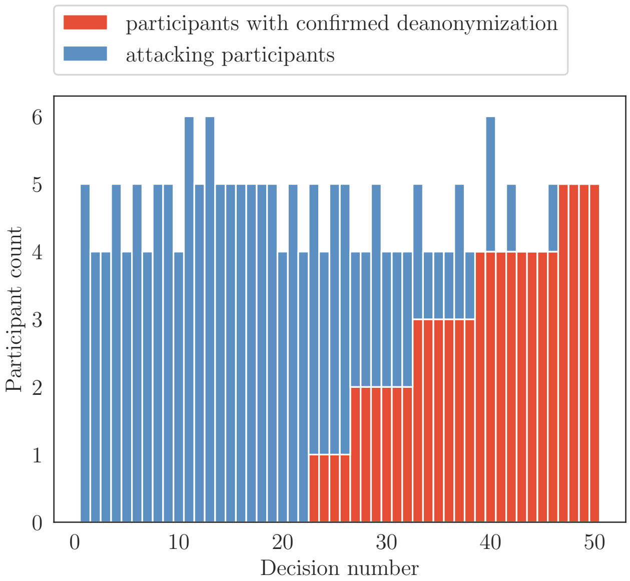


Figure 4: Attacking participants and participants with confirmed de-anonymizations per decision

7 Discussion

In this section, we discuss our findings, place them in the context of relevant related work and finally provide recommendations for anonymizing court decisions.

7.1 Which anonymized attributes or anonymization techniques are vulnerable? (RQ1)

We observed that all techniques and all attributes could be de-anonymized to some extent. This empirically confirms the increased risk of re-identification Opijnen et al. [40] suspected.

A rigorous statistical assessment of the precise impact or influence of each attribute or technique on de-anonymization, or a generalization of these findings to *all* decisions, is probably beyond what we can derive from our findings in Section 6.3. Indeed, anonymization techniques are not evenly distributed across attributes. Furthermore, confounding factors, such as the facts of a specific case, publicly available sources, or the skillset of the participants, may have played important roles in the de-anonymization process, and cannot be easily disentangled. As evidence of the potential impact of these external factors, some attacks managed to accurately de-anonymize non-preserving suppressions, which by definition preserve no information of the attribute-value string and thus should not be vulnerable. Another major factor that played a crucial role in de-anonymization was that many attribute-value strings were not anonymized at all, as we saw in our verification process: over 50% of the confirmed de-anonymizations were justified through unique inference.

With these caveats in mind, we can nevertheless glean important insights. First, non-preserving versions of the anonymization techniques seem less vulnerable than their partially preserving counterparts – which, from an information-theoretical perspective, makes sense. This is particularly evident for *omissions*: while partially-preserving and non-preserving omissions were roughly equally split in the set of attackable strings, partially-preserving omissions (initials) were almost four times more likely to feature in confirmed de-anonymizations than non-preserving (a priori) omissions (see Figure 1a). Nevertheless, lawsuits related to the insufficient anonymization of court decisions are still dismissed by some German courts on the grounds that initials offer sufficient anonymity [53, 54]. Our findings strongly suggest otherwise.

Second, among attributes, names of natural and legal persons as well as authorities accounted for over 50% of all confirmed de-anonymizations. This is especially problematic, since names directly identify the involved entities [4] that the anonymization process attempted to protect.

Recommendation 1: Non-preserving techniques Ideally, during writing, non-preserving (a priori) omissions should be used more extensively as they prevent attribute-value strings from arising. However, these omissions may indeed make it impossible to sustain information about relationships between different people, which would jeopardize decision comprehensibility. In such cases, non-preserving tagging or suppression should be employed, depending on how much information needs to be preserved for the decision to remain comprehensible. Pilán et al. [43] favor tagging over suppression. They argue that tagging better balances comprehensibility and privacy protection [43]. However, this may only be true under the assumption that both techniques protect data equally well, which our findings do not support. Furthermore, as long as the anonymization process remains mostly a manual effort in practice, suppression is arguably easier than finding suitable tags. Should anonymization be automated in the future, we recommend suppression over tagging. Tagging indeed requires suitable tags – a selection prone to error – while suppression might better protect the involved persons. Last, partially-preserving omissions should never be used under any circumstances.

Recommendation 2: Extensive anonymization In 2018, the Court of Justice of the European Union declared the removal of "any additional element likely to permit identification of the [natural] persons concerned" [17]. We agree and extend our recommendation to legal persons, as inferences can rapidly lead to de-anonymization. In short, one should not only anonymize relatively risky attributes (such as which type of athlete a person is), but also information that allows deriving such attributes (such as specific tournaments in which the person participated). To capture the relationship between these attributes and the additional information, the sanitization model of Sánchez and Batet [46] could help.

7.2 Which publicly available sources facilitate de-anonymizations? (RQ2)

Besides techniques, attributes and insufficient anonymization, publicly available sources also played a role in de-anonymization. Despite the large number of different source categories, news, personal pages and court decisions (including the attacked decision itself) remarkably accounted for 59% of all reported sources. While search engines were reported as a source in only 7% of the cases, our attackers spent a quarter of the total attack time using them. In other words, search engines are essential to help sift through information, but do not immediately provide information used in the actual de-anonymization.

Media sources play an important role as they accounted for 23% of all reported sources. Some of our decisions involved prominent people or attracted a lot of media attention (or both). This suggests that a fraction of all decisions will always be easy to de-anonymize, as already suspected by Nöhre regarding cases involving prominent people [36]. However, courts might deem the risk of de-anonymization justifiable, due to prominent persons being potentially entitled to less protection, and to increased legitimate public interest [12].

Other decisions and the attacked decision itself were 15% of all reported sources. In our verification process, they even accounted for one third of the sources. An analysis of the respective justifications showed that de-anonymization occurs in these cases because 1) the anonymized strings were either not anonymized or anonymized differently in decisions of lower or appellate courts, and/or 2) redactions of the attribute-value string were sometimes inconsistent in the attacked decision itself. Glaser, Schamberger, and Matthes [26] already pointed out these inconsistencies in anonymization practices; Csányi et al. [18] elaborated on the potential dangers.

Recommendation 3: Consistent anonymization First-instance decisions should already be anonymized in line with our previous recommendations, and identical standards should be held by subsequent courts. Such consistent anonymization could be achieved by nationwide guidelines. Consistency requirements may also affect the media. Journalists may have additional knowledge from, e.g., the trial, which they use in articles. Section 8 of the German Press Code [25] demands anonymization to be effective whenever required. The media should be guided by whether, what and how the courts anonymized. Thus, the media can ensure that they do not undermine what the courts deemed necessary to anonymize.

7.3 Which insights into reasonable effort can de-anonymization attacks provide? (RQ3)

Whether effort in terms of time, labor and costs is "reasonable" is a legal question for the courts to decide. However, the following empirical insights can support a legal assessment.

Time The average attack time was 33 minutes. Subtracting time spent in the application (e.g., for documentation), and the time spent in the decision, attackers only spent 16 minutes on publicly available sources. The majority of potential de-anonymizations occured in the first 18.5 minutes after the start of the attack. Many potential de-anonymizations were reported even after the time limit had expired (see Figure 3). Furthermore, less than 50% of all attackable strings were actually attacked. A qualitative evaluation of the participants' justifications revealed that they complained about being short on time and believed that they could find more information helpful for de-anonymizations with more time. In short, the limits on attack time we used were very stringent. In comparison, related work relied on longer attacks. Vokinger and Mühlematter [56] allowed researchers one hour per court decision to attack only two attributes (namely pharmaceutical companies and drugs). Tudor, Cornish, and Spicer [52] engaged authority employees for 3.5 hours for the de-anonymization of a statistical micro-dataset, while Branson et al. [8] allowed 24 hours to identify subjects from a medical dataset [8].

Labor The average participant was a 22.5-year-old law student in their sixth semester with experience in reading court decisions but not with data (de-)anonymization and who received no guidance from us. This is in line with a motivated intruder test, which relies on intruders without specialized knowledge or skills [27]. In comparison, Tudor, Cornish, and Spicer [52] used employees experienced in handling the data to be de-anonymized, and who received guidance on de-anonymization.

To account for different skill levels, *several* attackers should de-anonymize *one* decision. Employing at least two attackers makes sense as the optimum was 1.3 attackers, and the gains by choosing two over a single random attacker were largest.

As our verification process revealed that some potential de-anonymizations were incorrect, we inspected the participants' justifications and sources to find out why. We noticed that wrong conclusions were drawn due to a lack of legal knowledge and improper reading of the sources. These errors could potentially be prevented by providing de-anonymization guidelines and requiring more legal knowledge.

Costs Participants could only access publicly available information and had no budget at their disposal to access non-public sources. This is in line with a motivated intruder test, where only libraries, social media, internet searches, press archives and similar should be used [27]. German and European case law would potentially allow for more sources as they only require information to be *generally* accessible [39, 55] – e.g., public registers behind a paywall – or acquirable by legal means [11, 24]. Some participants stated in their justifications that they would have needed such an extended access.

In contrast to the work by Vokinger and Mühlematter [56], we did not use any technologies to facilitate or prepare de-anonymization. Our selection of decisions focused exclusively on their length and distribution of techniques and attributes. Participants were presented with the exact same decision available from the respective public portals.

Reasonableness To sum up, an average attacker that attacks a court decision in our experiment cost EUR 7 and took 33 minutes, during which they attacked 4.5 of 9.7 anonymized strings. 1.1 of those 4.5 attacked strings were confirmed de-anonymizations. Was the effort required to achieve this "reasonable?"

First, costs cannot be lower as only publicly available sources and no advanced technologies were used. Second, participant remuneration corresponds to the legal minimum wage at the time of writing, and is thus a lower bound. Third, the attack time was very short: participants reported that they needed more time. As corroborating evidence, they did not attack more than half of the attackable strings. All this implies that we worked with arguably *minimal* effort.

Assuming that such a minimal effort were deemed unreasonable from a legal standpoint, this would mean that almost everything would be legally considered anonymous. Furthermore, reasonable effort must be *more* than minimal effort as otherwise case law and legislators would use a "minimal effort" criterion to define anonymity. Thus, the effort in our experiment appears to be *reasonable* in the legal sense.

Recommendation 4: Conduct a crude adversarial evaluation of anonymization prior to public release In the process of anonymizing a court decision, crude de-anonymization attempts could already reveal obvious vulnerabilities, as our experiment showed. Even keeping in mind manpower shortage, personnel tasked with anonymization could work in pairs. Alice would anonymize a decision that Bob would quickly attempt to de-anonymize and provide Alice with feedback. Then, for a separate decision, roles would be reversed, with Bob anonymizing the decision and Alice performing the adversarial evaluation. As skilled professionals, Alice and Bob would likely be much faster and more effective than our participants, so that the imposition on their schedules should remain reasonable.

8 Conclusion

Our experiment with 54 law students showed *all* anonymization techniques and *all* anonymized attributes in 50 selected German court decisions were vulnerable to de-anonymization attacks, despite a restrictive attacker model that ensured only minimal effort could be expended, in terms of time, labor and cost. We argue that our conservative approach satisfies (and even exceeds) any plausible legal concept of "reasonable effort," which means that the supposedly protected information is not anonymous in a legal sense. This legal reasoning, in combination with the vulnerabilites we found, might help involved persons in future legal actions against the insufficient anonymization of court decisions. On a more positive note, we identified four practical recommendations that should greatly improve the situation, especially in a future where anonymization becomes automated as a means to address volume: using non-preserving techniques, performing extensive anonymization, using consistent anonymization techniques, and conducting crude adversarial evaluations before publishing court decisions.

Acknowledgments

We thank Axel Adrian, Kyle Crichton, Felix Freiling, Paul Gahman, Sarah Pearman, all anonymous reviewers and the shepherd of this work for their corrective feedback and support. Further, we thank the participants for their time and (reasonable) effort. Work was partially supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the Research and Training Group 2475 "Cybercrime and Forensic Computing" (grant number 393541319/GRK2475/1-2019).

References

[1] A. Adrian, S. Evert, M. Keuchen, P. Heinrich, and N. Dykes. "Anonymisierung von Gerichtsurteilen – Eine wesentliche Voraussetzung für E-Justice". In: *Jusletter IT* (May 2021).

[2] T. Allard, L. Béziaud, and S. Gambs. "Online publication of court records: circumventing the privacy-transparency trade-off". In: *1st International Workshop on Law and Machine Learning LML2020, in conjunction with ICML 2020*. Vienna, Austria, July 2020.

[3] Article 29 Data Protection Working Party. *Opinion 05/2014 on Anonymisation Techniques, 0829/14/EN WP 216*. Apr. 2014.

[4] Article 29 Data Protection Working Party. *Opinion 4/2007 on the concept of personal data, 01248/07/EN WP 136*. June 2007.

[5] M. Barbaro and T. Zeller. "A Face Is Exposed for AOL Searcher No. 4417749". In: *The New York Times* (Aug. 2006).

[6] D. Barth-Jones. *The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now*. Tech. rep. Rochester, NY: Social Science Research Network, July 2012.

[7] K. Benitez and B. Malin. “Evaluating re-identification risks with respect to the HIPAA privacy rule”. In: *Journal of the American Medical Informatics Association* 17.2 (Mar. 2010), pp. 169–177.

[8] J. Branson, N. Good, J.-W. Chen, W. Monge, C. Probst, and K. El Emam. “Evaluating the re-identification risk of a clinical study report anonymized under EMA Policy 0070 and Health Canada Regulations”. In: *Trials* 21.1 (Feb. 2020).

[9] Bundesgerichtshof (BGH) (Federal Court of Justice), 5 AR (Vs) 112/17, Jun. 20, 2018.

[10] Bundesgerichtshof (BGH) (Federal Court of Justice), IV AR(VZ) 2/16, Apr. 05, 2017.

[11] Bundesgerichtshof (BGH) (Federal Court of Justice), VI ZR 135/13, May 16, 2017.

[12] Bundesgerichtshof (BGH) (Federal Court of Justice), VI ZR 304/12, Nov. 05, 2013.

[13] Bundesverfassungsgericht (BVerfG) (Federal Constitutional Court), 1 BvR 1063/87, Sep. 28, 1987.

[14] Bundesverfassungsgericht (BVerfG) (Federal Constitutional Court), 1 BvR 857/15, Sep. 14, 2015.

[15] Bundesverwaltungsgericht (BVerwG) (Federal Administration Court), 6 C 3.96, Feb. 26, 1997.

[16] Council of the European Union. *Council Conclusions "Access to Justice – Seizing the Opportunities of Digitalisation" No 11599/20*. URL: https://data.consilium.europa.eu/doc/document/ST-11599-2020-INIT/en/pdf (visited on 10/08/2021).

[17] Court of Justice of the European Union. *Press Release No 96/18 "From 1 July 2018, requests for preliminary rulings involving natural persons will be anonymised"*. URL: https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-06/cp180096en.pdf (visited on 10/08/2021).

[18] G. M. Csányi, D. Nagy, R. Vági, J. P. Vadász, and T. Orosz. “Challenges and Open Problems of Legal Document Anonymization”. In: *Symmetry* 13.8 (2021).

[19] C. Culnane, B. I. P. Rubinstein, and V. Teague. *Health Data in an Open World*. 2017. arXiv: 1712.05627 [cs.CY].

[20] A. Datta, D. Sharma, and A. Sinha. “Provable de-anonymization of large datasets with sparse dimensions”. In: *International Conference on Principles of Security and Trust*. Springer. 2012, pp. 229–248.

[21] K. El Emam, E. Jonker, L. Arbuckle, and B. Malin. “A Systematic Review of Re-Identification Attacks on Health Data”. In: *PLOS ONE* 6.12 (Dec. 2011), pp. 1–12.

[22] European Commission for the Efficiency of Justice (CEPEJ). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*. Dec. 2018.

[23] A. Forget, S. Komanduri, A. Acquisti, N. Christin, L. F. Cranor, and R. Telang. *Security behavior observatory: Infrastructure for long-term monitoring of client machines*. Tech. rep. Carnegie Mellon University Pittsburgh United States, 2014.

[24] Gerichtshof der Europäischen Union (EuGH) (Court of Justice of the European Union), C 582/14, Oct. 19, 2016.

[25] German Press Council. *German Press Code Version of 22.03.2017*. 2017. URL: https://www.presserat.de/pressekodex.html (visited on 10/08/2021).

[26] I. Glaser, T. Schamberger, and F. Matthes. “Anonymization of German Legal Court Rulings”. In: *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*. ICAIL ’21. São Paulo, Brazil: Association for Computing Machinery, 2021, pp. 205–209.

[27] Information Commissioner’s Office. *Anonymisation: managing data protection risk code of practice*. 2021. URL: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf (visited on 10/08/2021).

[28] *Issue 391471: chrome.windows.onFocusChanged.-addListener is not fired*. URL: https://bugs.chromium.org/p/chromium/issues/detail?id=391471 (visited on 10/07/2021).

[29] V. Janmey and P. L. Elkin. “Re-Identification Risk in HIPAA De-Identified Datasets: The MVA Attack”. eng. In: *AMIA Annual Symposium proceedings* 2018 (2018), pp. 1329–1337.

[30] M. Keuchen and D. Deuber. “Öffentlich zugängliche Rechtsprechung für Legal Tech - Eine rechtliche und empirische Betrachtung im Lichte des DNG - Teil 2”. In: *Recht Digital* (2022), pp. 229–236.

[31] F. Linde and W. G. Stock. *Information Markets: A Strategic Guideline for the I-Commerce*. Berlin, New York: De Gruyter Saur, 2011.

[32] N. Mamede, J. Baptista, and F. Dias. “Automated anonymization of text documents”. In: *2016 IEEE Congress on Evolutionary Computation (CEC)*. 2016, pp. 1287–1294.

[33] G. Matthews and O. Harel. “Data confidentiality: A review of methods for statistical disclosure limitation and methods for assessing privacy”. In: *Stat. Surv.* 5 (Jan. 2011).

[34] A. Narayanan and V. Shmatikov. “Robust De-anonymization of Large Datasets (How To Break Anonymity of the Netflix Prize Dataset)”. In: *CoRR* abs/cs/0610105 (2006). arXiv: cs/0610105.

[35] A. Narayanan and V. Shmatikov. *Robust de-anonymization of large sparse datasets : a decade later*. 2019.

[36] I. Nöhre. “Anonymisierung und Neutralisierung von veröffentlichungswürdigen Gerichtsentscheidungen”. In: *Monatsschrift für Deutsches Recht* (Feb. 2019), pp. 136–141.

[37] L. O’Neill, F. Dexter, and N. Zhang. “The Risks to Patient Privacy from Publishing Data from Clinical Anesthesia Studies”. In: *Anesthesia & Analgesia* 122.6 (June 2016), pp. 2017–2027.

[38] Oberlandesgericht Frankfurt (OLG Frankfurt) (Higher Regional Court Frankfurt), 20 VA 21/17, Sep. 19, 2019.

[39] Oberlandesgericht Karslruhe (OLG Karlsruhe) (Higher Regional Court Karlsruhe), 6 VA 24/20, Dez. 22, 2020.

[40] M. Opijnen, G. Peruginelli, E. Kefali, and M. Palmirani. *On-Line Publication of Court Decisions in the EU: Report of the Policy Group of the Project Building on the European Case Law Identifier*. Jan. 2017.

[41] S. Pearman, J. Thomas, P. Emami-Naeini, H. Habib, L. Bauer, N. Christin, L. Cranor, S. Egelman, and A. Forget. “Let’s go in for a closer look: Observing passwords in their natural habitat”. In: *Proceedings of the 24th ACM Conference on Computer and Communications Security (CCS’17)*. Dallas, TX, Oct. 2017, pp. 295–310.

[42] A. Pfitzmann and M. Hansen. *A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management*. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf. v0.34. Aug. 2010.

[43] I. Pilán, P. Lison, L. Øvrelid, A. Papadopoulou, D. Sánchez, and M. Batet. “The Text Anonymization Benchmark (TAB): A Dedicated Corpus and Evaluation Framework for Text Anonymization”. In: *Computational Linguistics* (Aug. 2022), pp. 1–49. eprint: https://direct.mit.edu/coli/article-pdf/doi/10.1162/coli_a_00458/2040669/coli_a_00458.pdf.

[44] F. Reimer. “Juristische Texte lesen – Hilfestellungen aus öffentlich-rechtlicher Sicht”. In: *Zeitschrift für das Juristische Studium* (May 2012), pp. 623–629.

[45] L. Rocher, J. M. Hendrickx, and Y.-A. De Montjoye. “Estimating the success of re-identifications in incomplete datasets using generative models”. In: *Nature communications* 10.1 (2019), pp. 1–9.

[46] D. Sánchez and M. Batet. “C-sanitized: A privacy model for document redaction and sanitization: C-Sanitized: A Privacy Model for Document Redaction and Sanitization”. In: *Journal of the Association for Information Science and Technology* 67.1 (2016), pp. 148–163.

[47] D. Sánchez, S. Martínez, and J. Domingo-Ferrer. “Comment on “Unique in the shopping mall: On the reidentifiability of credit card metadata””. In: *Science* 351.6279 (Mar. 2016), pp. 1274–1274.

[48] L. Sweeney. “Simple Demographics Often Identify People Uniquely”. Working paper. 2000.

[49] L. Sweeney and J. S. Yoo. “De-anonymizing South Korean Resident Registration Numbers Shared in Prescription Data”. In: *Technology Science* (Sept. 2015).

[50] L. Sweeney, J. S. Yoo, L. Perovich, K. E. Boronow, P. Brown, and J. G. Brody. “Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study”. In: *Technology Science* (Aug. 2017).

[51] S. Trauzettel-Klosinski, K. Dietz, and the IReST Study Group. “Standardized Assessment of Reading Performance: The New International Reading Speed Texts IReST”. In: *Investigative Ophthalmology & Visual Science* 53.9 (Aug. 2012), pp. 5452–5461.

[52] C. Tudor, G. Cornish, and K. Spicer. “Intruder Testing on the 2011 UK Census: Providing Practical Evidence for Disclosure Protection”. In: *Journal of Privacy and Confidentiality* 5.2 (Feb. 2014).

[53] Verwaltungsgericht Berlin (VG Berlin) (Administrative Court Berlin), 27 L 43/20, Feb. 27, 2020.

[54] Verwaltungsgericht Meiningen (VG Meiningen) (Administrative Court Meiningen), 8 E 464/14 Me, Feb. 25, 2015.

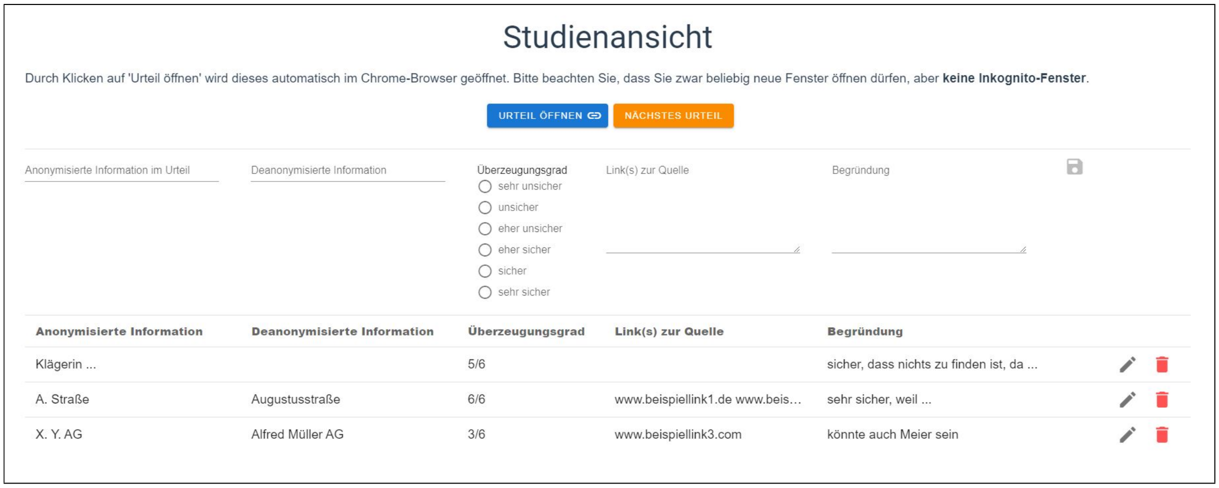
[55] Verwaltungsgerichtshof Mannheim (VGH Mannheim) (Higher Administrative Court Mannheim), 2 S 623/20, Jul. 10, 2020.

[56] K. Vokinger and U. Mühlematter. “Re-Identifikation von Gerichtsurteilen durch «Linkage» von Daten(banken)”. In: *Jusletter IT* (Sept. 2019).

[57] K. Vokinger, D. Stekhoven, and M. Krauthammer. “Lost in Anonymization - A Data Anonymization Reference Classification Merging Legal and Technical Considerations”. In: *J Law Med Ethics* (Mar. 2020), pp. 228–231.

A Application Interface

Figure 5 shows the interface of our application.



Studienansicht

Durch Klicken auf 'Urteil öffnen' wird dieses automatisch im Chrome-Browser geöffnet. Bitte beachten Sie, dass Sie zwar beliebig neue Fenster öffnen dürfen, aber **keine Inkognito-Fenster**.

Anonymisierte Information im Urteil

Deanonymisierte Information

Überzeugungsgrad
○ sehr unsicher
○ unsicher
○ eher unsicher
○ eher sicher
○ sicher
○ sehr sicher

Link(s) zur Quelle

Begründung

Anonymisierte Information	Deanonymisierte Information	Überzeugungsgrad	Link(s) zur Quelle	Begründung
Klägerin ...		5/6		sicher, dass nichts zu finden ist, da ...
A. Straße	Augustusstraße	6/6	www.beispiellink1.de www.beis...	sehr sicher, weil ...
X. Y. AG	Alfred Müller AG	3/6	www.beispiellink3.com	könnte auch Meier sein

Figure 5: Interface of the application used to document the results (all table entries are fictitious)

B Time Calculations

The times reported in Table 4 were calculated as follows. To determine whether the participant was in the browser or the application, we recorded, in the application, when the application window gained or lost focus. When an input or click occurred within the application, the application necessarily received focus; when an input or click happened in the browser, the application necessarily lost focus. As there was an unfixed (at the time) bug with the Chrome window events [28], we had to rely on the application events. Thus, application time is comprised of all the time the application had focus. Browser time is the total time spent on a decision minus the application time. As we displayed the decision in the browser, we were able to derive all other times (decision, Google and other publicly available sources) via Chrome events recorded by the SBO extension (see Section 4.4). Since participants had two screens to work with, the times presented do not necessarily coincide with the participants' actual attention time. The times should therefore not be understood as strictly separate but can overlap, for example, because the application was focused but the participant was reading in the browser.

C De-anonymization results

Table 5 shows the total number of attackable strings, attacked strings, potential and confirmed de-anonymizations broken down by technique and attribute. Figure 6 shows how many strings were attacked and how many of them were confirmed de-anonymizations for each decision.

Table 4: Average time spent per attack

Spent in	**Minutes** (Count)	(%)
Application	8.1	25
Decision	7.9	24
Google	7.4	23
Other sources	9.0	28
Total	*32.7*	*100*

(a) Premature termination included

Spent in	**Minutes** (Count)	(%)
Application	8.2	22
Decision	8.5	23
Google	8.8	24
Other sources	11.0	30
Total	*36.8*	*100*

(b) Premature termination excluded

Spent in	**Minutes** (Count)	(%)
Application	8.0	30
Decision	7.0	27
Google	5.2	20
Other sources	6.0	23
Total	*26.4*	*100*

(c) Premature termination only

Table 5: Total number of attackable strings, attacked strings, potential and confirmed de-anonymizations. (Key: AE: attackable strings; AD: attacked strings; PD: potential de-anonymizations; CD: confirmed de-anonymizations.)

	AE	**AD**	**PD**	**CD**
A priori omission, non-preserving	506	223	53	31
Omission, partially preserving	516	282	153	118
Suppression, non-preserving	650	277	71	40
Suppression, partially preserving	281	126	48	33
Tagging, non-preserving	204	87	33	20
Tagging, partially preserving	87	52	21	20
Total	*2 244*	*1 047*	*379*	*262*

(a) Broken down by technique

	AE	**AD**	**PD**	**CD**
Name (Natural person)	822	357	91	63
Name (Legal person)	464	242	82	47
Location (Street)	107	55	24	16
Location (City)	276	155	78	53
Location (Country)	40	15	10	10
Authority	124	54	27	24
Date	58	27	10	7
URL	152	72	35	24
ID	101	36	5	4
Miscellaneous	100	34	17	14
Total	*2 244*	*1 047*	*379*	*262*

(b) Broken down by attribute

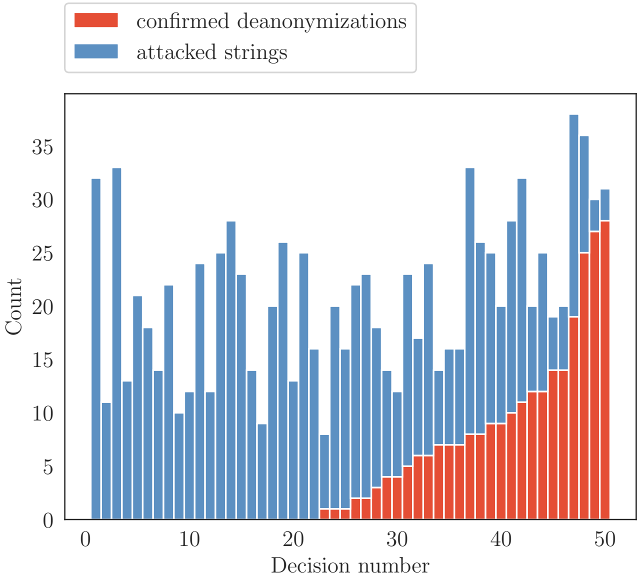


Figure 6: Attacked strings and confirmed de-anonymizations per decision