Minhui Xue, Gabriel Magno, Evandro Cunha, Virgilio Almeida, and Keith W. Ross*

The Right to be Forgotten in the Media: A Data-Driven Study

Abstract: Due to the recent "Right to be Forgotten" (RTBF) ruling, for queries about an individual, Google and other search engines now delist links to web pages that contain "inadequate, irrelevant or no longer relevant, or excessive" information about that individual. In this paper we take a data-driven approach to study the RTBF in the traditional media outlets, its consequences, and its susceptibility to inference attacks. First, we do a content analysis on 283 known delisted UK media pages, using both manual investigation and Latent Dirichlet Allocation (LDA). We find that the strongest topic themes are violent crime, road accidents, drugs, murder, prostitution, financial misconduct, and sexual assault. Informed by this content analysis, we then show how a third party can discover delisted URLs along with the requesters' names, thereby putting the efficacy of the RTBF for delisted media links in question. As a proof of concept, we perform an experiment that discovers two previously-unknown delisted URLs and their corresponding requesters. We also determine 80 requesters for the 283 known delisted media pages, and examine whether they suffer from the "Streisand effect," a phenomenon whereby an attempt to hide a piece of information has the unintended consequence of publicizing the information more widely. To measure the presence (or lack of presence) of a Streisand effect, we develop novel metrics and methodology based on Google Trends and Twitter data. Finally, we carry out a demographic analysis of the 80 known requesters. We hope the results and observations in this paper can inform lawmakers as they refine RTBF laws in the future.

Keywords: Privacy; Right to be Forgotten; Streisand effect; Latent Dirichlet Allocation

DOI 10.1515/popets-2016-0046
Received 2016-02-29; revised 2016-06-02; accepted 2016-06-02.

Minhui Xue: East China Normal University (ECNU) and NYU Shanghai, Email: minhuixue@nyu.edu;
Gabriel Magno: Federal University of Minas Gerais (UFMG), Email: magno@dcc.ufmg.br;
Evandro Cunha: Federal University of Minas Gerais (UFMG) and Leiden University, Email: evandrocunha@dcc.ufmg.br;
Virgilio Almeida: Federal University of Minas Gerais (UFMG) and Harvard University, Email: virgilio@dcc.ufmg.br;
***Corresponding Author: Keith W. Ross:** New York University (NYU) and NYU Shanghai, Email: keithwross@nyu.edu.

1 Introduction

In 2009 Mario Costeja González, a Spanish lawyer, requested that Google Spain remove a link to an online version of a 1998 article in the La Vanguardia newspaper about the required sale of his property due to social security debts. Because the sale had been concluded years before and the debt had been paid in full, he felt that information regarding his home-foreclosure notices was defamatory and no longer relevant. When the request was denied, Costeja sued Google. In May 2014, the European Court of Justice decided in favor for Costeja and ordered both Google Inc. and its subsidiary Google Spain to delist the pertinent links from Google's search results when querying Costeja's name. The court further ruled that search engines are required to remove from the list of search results, when requested by an individual, links to web pages that contain "inadequate, irrelevant or no longer relevant, or excessive" information about that individual. After this so-called "Right to be Forgotten" (RTBF) ruling, Google launched an online request process on May 29, 2014 and has since received more than 1.5 million link-removal requests from individuals ("requesters") in the European Union. It is important to note that the RTBF ruling does not effect the original published content; it only concerns the search-engine results (URLs) for specific queries.

The RTBF ruling has energized a debate concerning privacy on the Internet globally [1–3, 6, 8, 17, 18, 21, 22, 25, 28]. On one hand, it allows individuals to become directly involved in the process; and it provides a new way to minimize the damage associated with publicly inaccurate or outdated information. On the other hand, some argue that the RTBF restricts the right to freedom of speech. Many nations, and the United States in particular (with the First Amendment to the United States Constitution), have very strong domestic freedom of speech law, which would be difficult to reconcile with the RTBF.

The RTBF ruling also demonstrates the limits of national data privacy systems in a world of transnational data flows. The current RTBF ruling only applies to searches within European versions of Google (for example, within google.fr or google.uk); it does not apply to the US site google.com. Thus, when a search is made on google.com with the RTBF requester's name, the links to the RTBF-delisted content will continue to appear. Because google.com is still accessible to

any European, the French data protection authority CNIL has recently ordered Google to delist links from all of its geographic properties including google.com [10]. Google has so far refused, and the dispute is likely to end up in European courts.[1]

RTBF is one of the most important and controversial privacy laws of the decade, and will likely remain so for the years to come. To gain a deeper understanding of the RTBF ruling and its consequences, it is important to understand the type of content that is being delisted, the types of people making the requests, the consequences of delisting, and its potential susceptibility to data-driven inference attacks. As David Jordan, the head of editorial policy at BBC said concerning the RTBF [5], "*It's impossible to have a meaningful debate if you've not got an idea about what's being delisted.*"

In this paper we take a data-driven approach to study the RTBF *as it applies to traditional media outlets*, such as newspapers and broadcasters. Several traditional media sources in the UK have republished links that were delisted by Google. We collect 283 such republished delisted links and perform an extensive analysis on their collective content. We examine the following research questions:

- **What types of incidents or events are being delisted?** To answer this question, we take two approaches. First we manually classify each of the 283 delisted articles into a number of categories. Second, we apply Latent Dirichlet Allocation (LDA) to automatically and objectively determine the topics of the delisted content, and automatically classify the articles among the topics.
- **Is it possible for a third-party to automatically determine delisted links, even if the links have not been republished?** We discuss a data-driven inference attack that "transparency activists" or other third parties could take to discover delisted links. Such groups could then post the rediscovered links on their web sites, along with the names of the requesters, thereby undermining the goals of the RTBF ruling. We demonstrate the feasibility of the attack with an experiment that discovers two previously-unknown delisted articles.
- **Are some RTBF requesters experiencing a Streisand effect?** The "Streisand effect" is the phenomenon whereby an attempt to hide, remove, or censor a piece of information has the unintended consequence of publicizing the information more widely, usually facilitated by the Internet [9]. Costeja himself suffers from the Streisand effect – although he won this landmark case, it is unlikely he will ever be forgotten because his name now appears on thousands of web sites. We explore whether the republishing of delisted links can engender the Streisand effect. To measure the presence (or lack of presence) of a Streisand effect, we develop novel metrics and methodology based on Google Trends and Twitter data.
- **What types of people are making the requests for delisting of media articles?** We are able to determine the identities of 80 individuals who made RTBF requests for 103 out of the 283 articles. Acknowledging that all of the requests are for UK media sites, we perform a demographic analysis on these 80 individuals, gaining insight into their gender, age, professions, and celebrity.

To the best of our knowledge, this is the first data-driven study of RTBF based on actual delisted content. The results and observations in this paper can inform lawmakers as they refine the RTBF laws in the future, and in particular, inform the current debate on whether the RTBF law should be extended beyond the European Google sites to all of Google's sites.

This paper is organized as follows. In Section 2, we provide overview of the RTBF procedure and also provide some basic RTBF statistics. In Section 3, we show our data set. In Section 4, we perform an analysis of RTBF delisted content for traditional media sources, including an automated analysis based on LDA. In Section 5, we discuss possible schemes to automatically determine delisted RTBF links. In Section 6, we analyze the presence (or lack of presence) of a Streisand effect for the 80 distinct requesters. In Section 7, we perform a demographic analysis on 80 distinct requesters. In Section 8, we survey the related work. In Section 9, we conclude with some recommendations.

1.1 A Note on Ethics and Privacy

The goal of this paper is to gain deeper insights into the consequences to the RTBF ruling as it applies to traditional media content, and understand possible attacks that could be eventually launched against it. We note that the data analyzed in this paper is publicly available and easily obtained. From this publicly-available data we identify 80 requesters. We take precautions in this paper not to mention the names of these individuals, and we will not be providing the names of these individuals to anyone outside of our author list. Also we demonstrate the feasibility of uncovering delisted links with an experiment, and identify two previously-unknown delisted articles. Again, we will not provide the title of the article or the name

1 Very recently Google has agreed to also delist the links on google.com for searches made from within the European Union [24, 26]. However, the links continue to appear on google.com when searching from a physical location outside the European Union. Europeans in Europe can therefore use proxy services to see the links on google.com.

of the requester. Finally, we have alerted Google about the discoveries made in this paper.

2 Overview of the Right to be Forgotten

In this section we provide an overview of how Google manages the RTBF process. If an individual wants to request that a particular link be delisted from the European Google sites, the individual must first complete a web form provided by Google. Google also allows people to make requests on behalf of others, so long as they can affirm that they are legally authorized to do so. Since Google began accepting requests on May 29, 2014, Google has received approximately $429,000$ requests for the removal of approximately 1.5 million URLs as of May, 2016.

A committee at Google assesses each request on a case-by-case basis to determine whether the URLs should be removed or not. URLs are typically removed for the following types of requests [13]:

- Private or sensitive information, such as pages that contain information about personal contact, address, health, sexual orientation, race, ethnicity, and religion.
- Content that relates to minors or to minor crimes that occurred when the requester was a minor.
- Acquittals, exonerations, and spent convictions for crimes. Google tends to delist content relating to a conviction that is spent or accusations that are proven false in a court of law.

Google may decline to delist if it determines that the page contains information which is in the public interest. Determining whether content is in the public interest depends on diverse factors, including – but not limited to – whether the content relates to the requester's professional life, a past crime, political office, position in public life, or whether the content itself is self-authored content, government documents, or journalistic in nature [13]. As of May 2016, Google has removed approximately 43% of the requested URLs [14].

URLs are only delisted in response to queries relating to an individual's name. So, if Google grants a request to delist an article for John Doe about his trip to Shanghai, Google would not show the URL for queries relating to [john doe] but would show it for a query like [trip to shanghai]. This delisting policy is central to the attacks described in this paper.

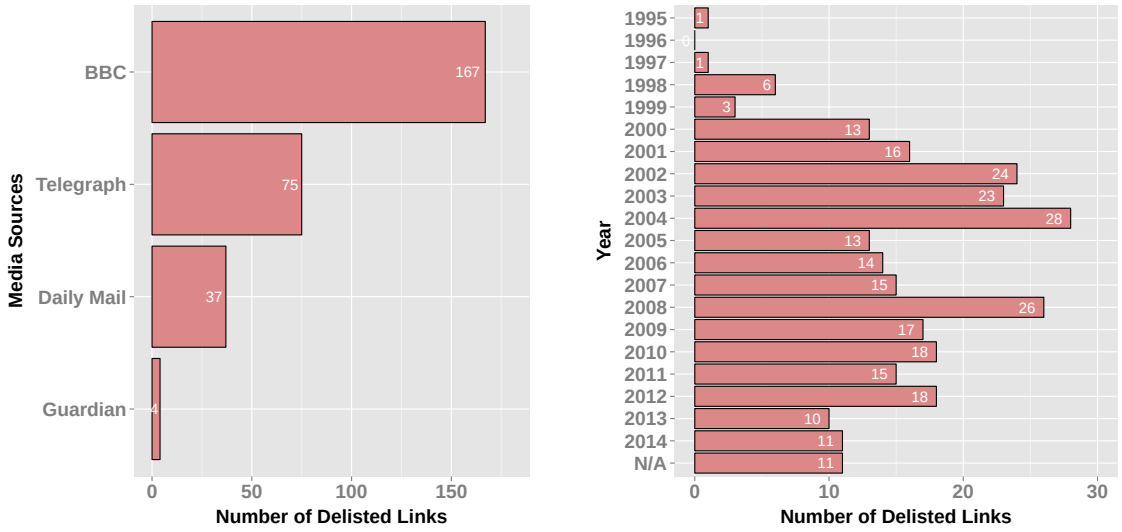
Google notifies webmasters when pages from the webmasters' sites are delisted. "In order to respect the privacy of the individuals who have made removal requests, Google only sends the affected URLs, not the requester's name [13]." For this paper, this particular action taken by Google will play an important role for our data collection and analysis.

Much of the discussion around the RTBF, including the Mario Costeja González court case, has been around online-media content that relates to legal issues, crimes, politicians and public figures. However, perhaps surprising to the architects of the law, the majority of the 1.5 million URL removal requests to date are for pages on social media and profiling sites that contain private personal information such as email address, home address, health, sexual orientation, race, ethnicity, religion, and political affiliation. In particular, each of the eight sites for which Google receives the most requests are either social media or profiling sites [14], and 95% of the requests are for delisting of URLs pointing to private information [27]. In this paper we study the requests that are made for content on mass media sites. Although the fraction of such requests is relatively small, in absolute numbers it is significant (approximately $75,000$ requested URLs) and, arguably, concerns the most controversial RTBF requests and content.

3 Data Set

As discussed in the previous section, Google has chosen to notify webmasters when pages from webmasters' sites are delisted. A number of media sites in the UK (oddly uniquely in the UK), upon receiving these notifications, republish the URLs in the name of transparency and full disclosure. For example, the BBC most recently republished links on June 25, 2015 [4]. As of December 2015, the BBC, the Telegraph, the Daily Mail, and the Guardian have republished a total of 283 delisted links. Of course, the web pages for these delisted links continue to exist on the web sites for the BBC, Telegraph, Daily Mail, and Guardian.

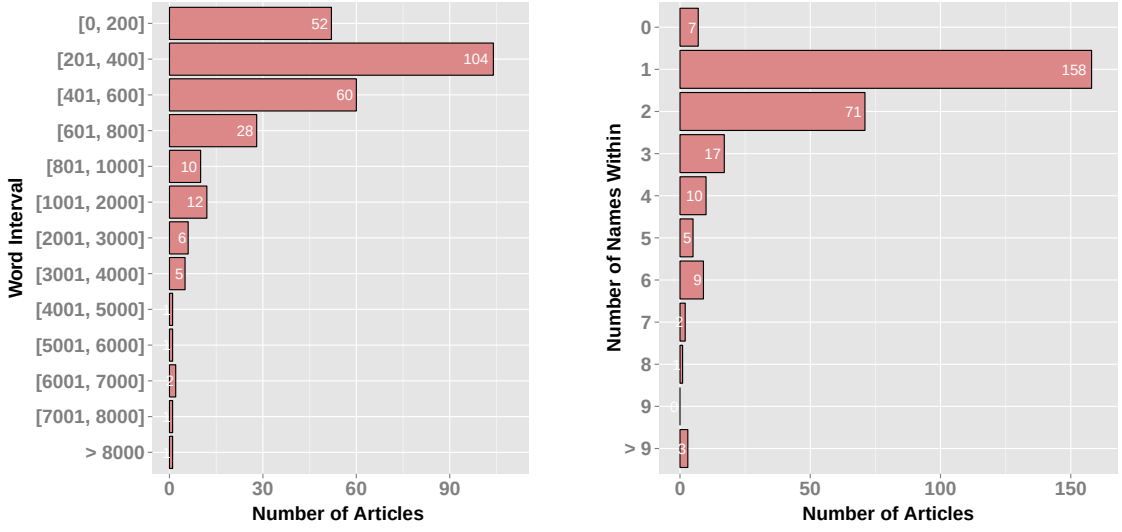
For each of the 283 delisted links, we downloaded the corresponding articles. These 283 articles constitute our basic data set. Figure 1 and Figure 2 show the basic properties of our underlying data set. Figure 1(a) shows that the BBC (59.0%) and the Telegraph (26.5%) take up more than three-quarters of the articles in our data set. From Figure 1(b), we see the dates of publication of these articles range from 1995 to 2014, with the large majority appearing between 2000 and 2012. Figure 2(a) shows that the articles greatly vary in size, with the minimum, maximum, and median number of words in an article given by 18, 15009, 370. The mean is given by 680.7 and the standard deviation is by 1282.9. We used the



(a) Distribution of delisted links across media sources

(b) Distribution of delisted links by year

Fig. 1. Basic Properties of the Data Set: Media Sources and Publication Dates



(a) Distribution of articles by number of words

(b) Distribution of articles by number of names within

Fig. 2. Basic Properties of the Data Set: Textual Properties

Stanford NER tool[2] to extract the names from the 283 articles. Figure 2(b) shows the distribution by the number of names within the articles. The minimum, maximum, median, mean, and standard deviation are given by 0, 52, 2, 2.1, and 2.1. As discussed in Section 5.1, these extracted names are candidate RTBF requesters. Note the existence of seven articles with zero names, making it difficult to determine candidate requesters for these delistings.

4 Analysis of Forgotten Media Content

In order to have meaningful debate about the RTBF, it is desirable to have a sense of the type of media content that is

2 http://nlp.stanford.edu/software/CRF-NER.shtml

being delisted in search engines. In this section, we take two approaches to content analysis: a manual approach; and an automated topic-analysis approach.

4.1 Manual Topic Analysis

We read all of the 283 articles and manually classify them to 18 different categories. Figure 3 shows the distribution of the 283 articles by categories. We see that there are four topics related to sexuality, which are "Sexual Assault," "Prostitution," "Pedophilia" (typically involving interactions between adults and minors), and "Sexual Miscellaneous." (If the article discusses a sexual incident but has nothing to do with assault, prostitution, or pedophilia, we then categorize it into "Sexual Miscellaneous.") Articles in "General Miscellaneous" are largely mundane topics related to sports, education, and so on. "Non-textual" consists of non-textual documents such as images. In Figure 3, the 18 categories are shown in decreasing order in terms of number of articles.

In general, we see that many of the delisted topics treat highly sensitivity topics, including sexuality, sexual assault, murder, pedophilia, financial misconduct, terrorism, and so on. Most likely Google accepted to delist many of these sensitive articles due to spent convictions, accusations that are proven false in a court of law, or content relating to a criminal charge for which the requester was acquitted [13]. Although not carried out in this paper, one could try to determine why a given article was delisted, perhaps by looking for other related articles, or by examining public or judicial records, discussing the same incident. Although this would be a fairly laborious and painstaking task for the 283 articles, it could be of interest to lawmakers in the future.

4.2 Automated and Objective Content Analysis

We also use Latent Dirichlet Allocation (LDA) to analyze the content in the 283 delisted articles ("documents" in LDA terminology). The goal of LDA is to study observable word occurrences in the articles and to determine the latent article-topic and topic-word distributions that help determine the themes discussed in each article [7]. The inputs to LDA consist of (i) the articles, (ii) the number of desired topics, (iii) and some distribution parameters. The output consists of the latent topics, with each topic defined as a distribution over the collection of all words in the corpus; and for each article, a distribution over the set of all topics.

Before performing LDA, we first pre-process the data. In order to avoid statistical outliers, we sanitize each document

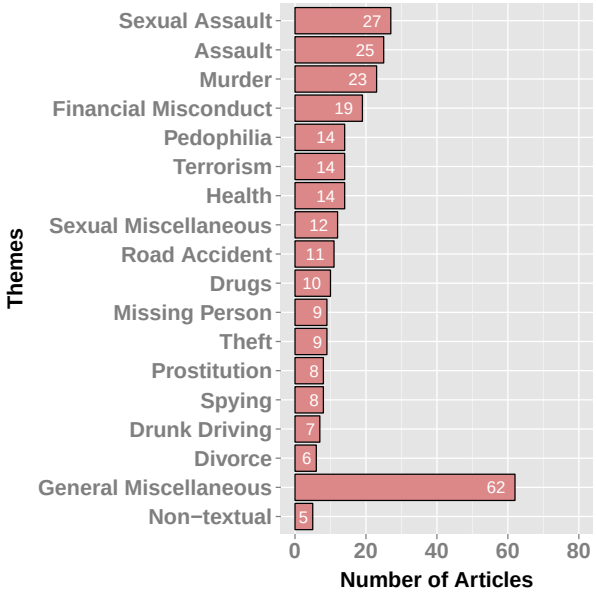
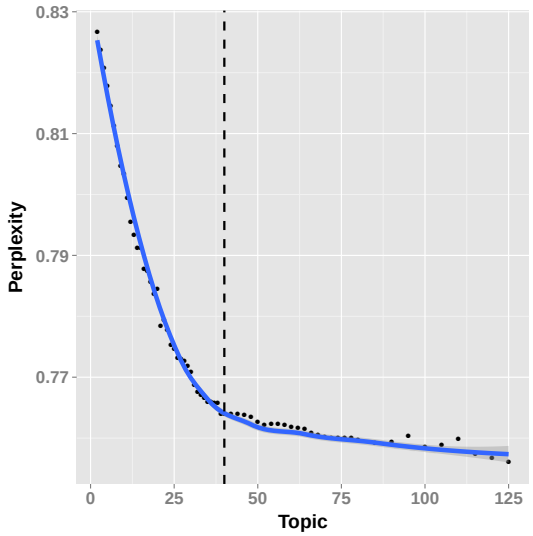


Fig. 3. Distribution of articles by theme

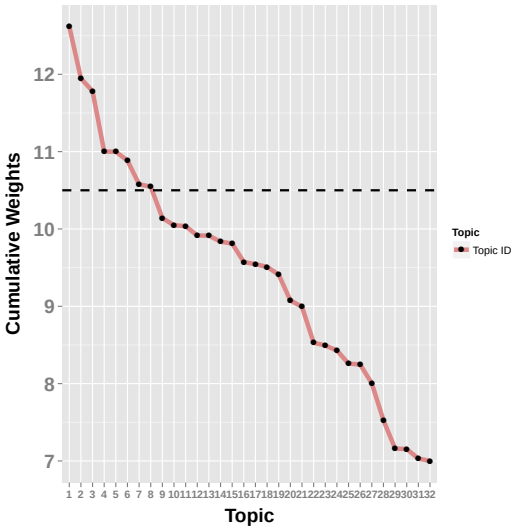
by filtering out common stopwords, determiners, verbs, numbers, proper nouns, and low frequency words that appear in less than 1% of all the forgotten content. We then perform the LDA topic analysis on these 283 documents using the Stanford Topic Modeling Toolbox.[3] When using LDA, we need to take as input the number of topics and Dirichlet prior probabilities for the topic-word and document-topic distributions. We then perform a *Perplexity* analysis to choose the number of topics. Based on our analysis shown in Figure 4(a), we choose 32 topics for our study. We also choose Symmetric Dirichlet priors (*i.e.*, 1/number of topics) by default for topic-word and document-topic distributions based on performance. In summary, we perform the LDA analysis on the 283 articles with 32 topics and Symmetric Dirichlet priors by default as the parameters.

For each identified topic, we then calculate the cumulative weight for that topic across all the documents. As illustrated in Figure 4(b), 32 topics are ordered by decreasing value of the cumulative weight. The eight topics with the highest cumulative weights along with the top five terms occurring in each topic are shown in Table 1. As shown in Table 1, the strongest topic themes are violent crime, drunk/drugged driving, domestic drug use, murder, prostitution, financial misconduct, and sexual assault. Other striking topics, such as health and spying, are also in the top half (but not shown in Table 1). We also show which documents are most relevant to the themes, helping us to quickly access the documents to validate the

3 http://nlp.stanford.edu/software/tmt/tmt-0.4/



(a) LDA: Perplexity Performance



(b) LDA: Cumulative Weights for Topics over All the Documents

Fig. 4. LDA Performance Analysis

Topic ID (Rank)	Cumulative Weight	Top Terms	Most Relevant Articles (Document ID)	LDA Theme
1	12.62	indecent, image, suspect, murder, inspector	4, 6, 89, 92, 93, 142 195, 226, 255, 268, 271	Violent Crime
2	11.95	drive, smash, car, cash, drug	25, 45, 49, 52, 133 149, 184, 220, 221	Drunk/Drugged Driving
3	11.78	debate, referee, heroin, wife, inject	1, 98, 103, 128, 213 233, 234, 245, 248, 265	Domestic Drug Use
4	11.00	jail, murder, corruption, convict, prosecution	96, 101, 113, 114, 123 136, 183, 196, 238, 250	Murder
5	11.00	passenger, escort, seat, benefit, solicitor	24, 63, 71, 164, 172 180, 190, 209, 239, 261	Prostitution
6	10.88	wife, bank, transfer, money, reassure	42, 70, 94, 100 151, 167, 207, 247	Financial Misconduct
7	10.58	woman, bed, charge, explosive, rape	13, 41, 46, 65, 83 87, 91, 223, 256	Sexual Assault
8	10.55	sex, rape, girl, deny, flat	57, 120, 159 162, 212, 270	Sexual Assault

Table 1. LDA Topics for Delisted Content

results. Comparatively, the LDA approach not only analytically validates the results from the manual approach applied in Section 4.1, but also further categorizes the topics previously belonging to "General Miscellaneous" into other categories, leading to a fine-grained classification.

5 Attacking the Right to be Forgotten

Informed by the data analysis of the previous section and additional trial-and-error experimentation, we now show how a third-party (such as a transparency activist [15] or a private

investigator) can "attack" the RTBF. First we show, given a known delisted article, how a third-party can often determine the name of the person who made the request to delist the article (the "requester"). We apply this first attack to our data set of 283 articles and determine the requesters for 103 articles. Second, we describe how a third-party attacker can rediscover delisted links. For this attack, we run an experiment as a proof of concept and uncover two previously-unknown delisted links. The attacker could combine these two attacks to determine large swaths of delisted links and their corresponding requesters, and then publish the requesters' names and the corresponding delisted links on a web site (such as [15]), thereby putting the efficacy of RTBF for media links in question. The purpose of this section is to alert the privacy community, government regulators, and operators of search engines to the possibility of these attacks.

5.1 Identifying the Requesters

Given a known RTBF delisted page, how can a third-party determine who made the request for delisting? Such a determination can be considered a privacy breach that undermines the spirit of the RTBF ruling. Indeed, Google writes on its FAQ page, "In order to respect the privacy of the individuals who have made removal requests, Google only sends [to the webmasters of the delisted pages] the affected URLs, not the requester's name [13]." We now provide an attack that can often determine the requester of a given delisted link.

Recall that Google does not display the delisted link whenever a query is made that includes the requester's name. The attack is based on the following simple observation: *In many cases, we would expect the requester to be mentioned in the article*. Given that the mean number of names in an article is 2.1 (see Figure 2(b)), the number of candidate requesters for any one delisted article is typically small. In the following attack, we use google.uk for concreteness; but the attack applies to all country extensions, such as google.es and google.fr.

Find Requester Attack:

Step 1. For each name in the article, put into google.uk the search query ["*name*," "*title*"] where *name* is the full name of the candidate and *title* is the article title.

Step 2. If google.uk returns the delisted URL (typically the first link due to the specificity of the request), then we conclude that *name* is not the requester.

Step 3. However, if it doesn't return the URL, then we conclude that *name* is the requester, and we refer to *name* as a *verified requester*.

We test this attack on the 283 articles. Specifically, using the Stanford NER tool,[4] we automatically extract the names from the 283 articles. We put each of these requester names along with the title of the article, as just described, into google.uk and discover 80 requesters for 103 delisted links. For the remaining 180 delisted links, the algorithm failed to identify the requesters. For those 180 articles, it is very likely that the requester is not actually listed in the article, but the requester nevertheless feels that the article compromises his/her privacy or reputation. A more sophisticated attack, involving examining related articles and/or public records, could potentially generate additional candidate requester names (which could be verified using the algorithm just described). We note that we also tried using companies and other entity names from the articles, and names from the comments for the articles, but this did not generate any additional requesters.

To double-check that a verified requester is indeed a requester, we can also put ["*name*," "*title*"] into google.com. We found that in *every case* where the delisted link does not appear on google.uk, it does appear on google.com. We emphasize that this verification involving google.com is only a validation step and is *not* necessary for the attack; thus, if the RTBF links were to be delisted from google.com (as well as from the European Google sites), the attack would still be valid.

5.1.1 How Difficult is it to Rediscover Forgotten Content using google.com?

The French data protection authority CNIL has recently ordered Google to delist links from all of its geographic properties including google.com [10]. Google has so far refused, and the dispute is likely to end up in European courts.[1] To inform this debate, we now attempt to provide some quantitative insights into this issue. Specifically, when querying with any one of the 80 known requester names, we investigate whether the corresponding delisted link (*i.e.*, delisted in Europe) appears on google.com, and if so, how deep in the search results it appears. The results are given in Figure 5. We see that for more than half of the requester names, the desired forgotten link can be rediscovered within the first 20 search results. Although for 26 requesters the corresponding forgotten links do not appear in the top 200 results, they may appear deeper in the results, perhaps in the thousands. For these 26 names, when we query with ["*name*," "*article title*"], the corresponding links always appear in the first ten results. We can conclude from this analysis that for the majority of the requesters (more than 50%), it is easy for a third-party to find the forgotten content

4 http://nlp.stanford.edu/software/CRF-NER.shtml

in google.com; however, for a fraction of users, the forgotten content only exists deep in the search results and requires significant effort to rediscover.

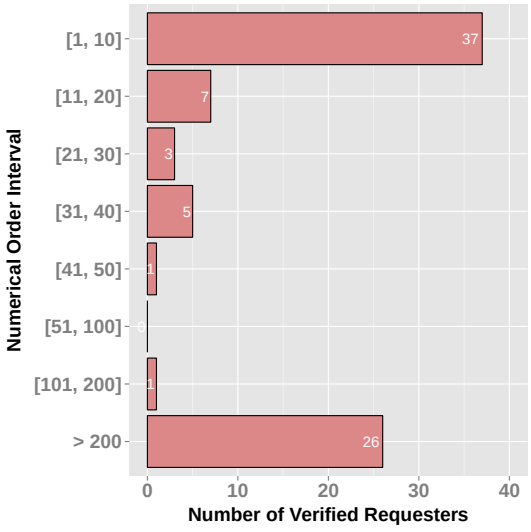


Fig. 5. Depth of delisted link for the 80 verified requesters

5.2 Rediscovering Delisted URLs

We now consider how a third-party can automatically find large swaths of delisted links, thereby putting the efficacy of the RTBF in question. In this attack, the attacker would take a divide and conquer approach, searching for delisted URLs for one target media source at a time. The idea is to first identify articles that are candidates for delisting directly from the target media source; then for each of the candidate articles, check to see it is indeed being delisted. We emphasize that this attack does *not* make use of google.com; the attack would be valid and unchanged even if the RTBF law were fully extended to google.com. To make the attack concrete, let us assume the attacker is attempting to find delisted links for the Spanish newspaper, El Mundo.

Find Delisted URLs Attack:

Step 1. Crawl and download all the digital articles of a media source such as El Mundo.

Step 2. Put all the articles in a database and automatically search on terms related to topics typically requested for removal, as given in Section 4, such as financial fraud and sexual abuse.

Step 3. From those articles, automatically collect all the names mentioned in those articles.

Step 4. Write a script that puts each name and article title into google.es and checks the links returned. As described in Section 5.1, if for any of these queries the link to the article is not in (the top ten) search results, then the third party has identified a delisted URL and the corresponding requester.

From the attacker's perspective, we now briefly analyze this attack. The attacker first needs to collect pages from the media source. Some media sources provide their articles on DVDs, and others provide search features which allow the attacker to find and download only the articles relevant to the desired RTBF search terms. Steps 2 and 3 can be easily and quickly done with automated tools. The effort required to carry out Step 4 depends on the number of candidate articles obtained in Step 2. Since there are on average 2.1 names in each article (see Section 3), if C is the number of candidate articles, then the expected number of search queries is $2.1 \times C$. Search engines employ rate limiters, preventing bots from querying from the same location (determined by IP address and other factors) too frequently. On the other hand, there are means to circumvent rate limitation by employing VPNs, proxies, botnets or cloud services to gain access to many IP addresses [20].

As a proof of concept, we run an experiment to demonstrate the feasibility of the attack. We first crawl the El Mundo web site (http://www.elmundo.es) with the goal of retrieving potential articles being delisted because of the RTBF law. We compiled a list of 37 terms in Spanish related to crime and investigations, such as "crimen," "fraude," and "abogado." Then, we queried the terms in the El Mundo web site and collected all URLs returned. In total we collected $85,123$ unique URLs for El Mundo articles. After that, we retrieved the HTML content corresponding to the URLs. Then, we ran the Stanford NER tool with a Spanish model to extract the names of people mentioned in the articles. In total, this generated a total of $259,812$ combinations of article + names. In order to reduce the number of queries for the experiment, we filter articles and names that we believe are more prone to be a RTBF request. First, in order to filter out popular names (*e.g.*, "barack obama," "mariano rajoy"), we ignore names (and associated queries) that appear in more than three articles. Also, to reduce the impact of very popular names and select more specific names, we only consider names with at least three "chunks," *e.g.*, "josé maría viñals," "josé castillo sánchez." Since El Mundo publishes news about events all over the world, we filter out articles about non-Spanish events; to this end, we only used URLs that match "http://www.elmundo.es/elmundo/XXX" (ignoring URLs like "http://www.elmundo.es/america/XXX"). Finally, we keep only articles that have one of these three words in its content: "condenado," "proceso," and "fraude." With all this filtering, we reduce the number of candidate articles to $4,164$ and the number of combinations of article + names to $6,410$. We then sequentially send these queries to google.es.

From the query results, we find two (unrelated) articles that meet the conditions of Step 4. We then manually investigated the articles and the names, verifying that the google.es lists the articles when querying only by article title, but does not list the articles when querying on the person's name or querying by the person's name along with the article title, providing strong evidence that the articles are being delisted because of the RTBF law. *In summary, in this experiment, from* $4,164$ *candidate El Mundo articles, we were able to find two previously-unknown RTBF delisted articles along with the requester names for these articles.* To scale this attack, the attacker would need to circumvent Google's query rate limitation mechanisms. We conjecture that an attacker with modest resources could employ a blackmarket botnet to scale this attack to explore not only all of the El Mundo articles, but all of the articles in most of the major European newspapers.

We emphasize that this attack will not uncover *all* of the delisted URLs for El Mundo, even if we consider every El Mundo article as a candidate article (no filtering). This is because the requester's name does not always appear in the article, as described in Section 5.1. Indeed, in Section 5.1, we only determine the requesters for 103 of 283 UK articles. But given the results with the 283 UK articles, we conjecture that this attack can potentially determine $30\%-40\%$ of the delisted media URLs across the European Union.

Are there any measures Google could take to prevent this attack? The attack works because the Google search results are different when the attack queries with and without the requester's name; specifically, the article is only delisted when the query includes the requester's name. A defense to this attack would be for Google to *always* delist the article, whether or not the query includes the requester's name. But many people would consider this to be a strong form of censorship, going significantly beyond the original intent of the RTBF.[5] Moreover, if the attacker knows that an article exists (by crawling a media site), and sees that the article is never listed for any query, then attacker can suspect that the article was delisted because of the RTBF.

Can this attack be extended beyond media sites to other sites such as social networks or profiling sites? As with El Mundo, an attacker can target a specific site such as profileengine.com, download the site's pages, extract names and information provided on the site about the names (such as birthdates and addresses), and then using each name and its associated information, query google.es one name at a time. Assuming the information is specific enough, if the link is not hidden, then it should appear among the top search results. But if the link does not appear among the search results, then the link is likely being delisted by Google and the name is the corresponding requester. More work is needed to study how much page collection and query effort would be needed to carry out the attack for different target sites. We believe, however, that it will be significantly more difficult to apply this approach for many of the non-media sites. For example, according to the Google Transparency Report [14], Facebook is the site for which the number of delisted URLs is the largest, with approximately $13,000$ delisted links as of May 2016. But Facebook has over 1.65 billion active users, with all of these users having a publicly available profile page, and many of these users providing large amounts of publicly available content (photos and text). Finding a relatively small number of delisted links in this ocean of content will likely be quite challenging.

6 Streisand Effect

The "Streisand effect" is the phenomenon whereby an attempt to hide, remove, or censor a piece of information has the unintended consequence of publicizing the information more widely, usually facilitated by the Internet [9]. Costeja himself suffers from the Streisand effect – although he won this landmark case, it is unlikely he will ever be forgotten because his name now appears on thousands of web sites. In this section we explore whether the republishing of delisted links by traditional media sources can engender the Streisand effect. To measure the presence (or lack of presence) of a Streisand effect, we develop novel metrics and methodology based on Google Trends and Twitter data.

6.1 Google Trends and Twitter

We put the names of each of the identified requesters into Google Trends[6] to obtain the relative number of queries for two months before and two months after republication of the corresponding delisted URL. Google Trends is a platform where one can see how often a particular search-term was queried in Google Search over time. It is possible to query single terms or compare the popularity of multiple terms. The value returned by Google's API is not the absolute number of queries, but an abstract value that represents the relative amount of queries for some particular time frame. We use a

5 Consider an article for which two people are arrested for sexual assault, but after the publication of the article, one person (the requester) is acquitted. Should this article never appear in the search results, no matter what the query?

6 https://www.google.com/trends/

Web Site	# of names	# of names Pos. gain	# of names Neg. gain	# of names Inf. Pos. gain	# of names Inf. Neg. gain	Avg. Pos. gain	Avg. Neg. gain	Max. Pos. gain	Min. Neg. gain
Google	22	10	12	3	4	1.69	−1.35	3.08	−2.76
Twitter	44	20	24	10	6	9.31	−2.34	31.48	−6.18

Table 2. Visibility Statistics

value that is normalized considering a time period of one year. Also, we use global Google Trends data rather than geographic specific Google Trends data.

For each of the 80 requesters, we also put their names into Twitter and collected all the tweets that mention the name for two months before and two months after republication. To accomplish this task, we used Twitter's "advanced search" feature,[7] which allows one to search for tweets with certain terms from a specific date or within a range of dates. Twitter shows results from the full index in the "all" tab of the search results, sorted by date. Unfortunately, Twitter does not provide an API to query the full index, so we had to collect tweets by issuing HTTP requests to the Twitter web server. Since we are interested only in the frequency of tweets instead of the actual content or any other metadata, we parsed and stored only the timestamp of the tweets. The query terms used were the full names of the requesters enclosed by quotation marks, meaning that we were interested in tweets mentioning the exact string name. We limited our search to only tweets of two months before and two months after republication.

6.2 Metrics: Quantifying Visibility

In order to analyze whether there is a Streisand effect, we first defined a *visibility* metric. The visibility $V_{(r,d,m)}$ depends on a particular requester r, a date (day) d, and a media m, and is defined as:

$$V_{(r,d,\text{Google})} = \text{relative \# of searches to } r \text{ on day } d;$$
$$V_{(r,d,\text{Twitter})} = \text{\# of tweets with } r \text{ on day } d.$$

We define the visibility to be zero when there is no data available (*e.g.*, no tweets mentioning a requester in a particular day, etc.).

For each delisted link, we obtain data for two months before and two months after the date of republishing. Specifically, for each requester r we also denote the date of republication D_r of the delisted link. We also define the average of the visibility before (Mb) and the average of the visibility after (Ma) the D_r for the fixed time frame $T_f = 60$ days:

$$Mb_{(r,m)} = \frac{1}{T_f} \cdot \sum_{d=D_r-61}^{D_r-1} V_{(r,d,m)};$$
$$Ma_{(r,m)} = \frac{1}{T_f} \cdot \sum_{d=D_r}^{D_r+60} V_{(r,d,m)}.$$

Finally, we define the *gain* G for a requester r and a media m as follows:

$$G_{(r,m)} = \begin{cases} \frac{Ma_{(r,m)}}{Mb_{(r,m)}}, & \text{if } Ma > Mb \\ -\frac{Mb_{(r,m)}}{Ma_{(r,m)}}, & \text{if } Mb \geqslant Ma \end{cases}$$

The gain $G_{(r,m)}$ is to be interpreted as the change of visibility that a requester r has in media m after the republication of the corresponding delisted link. A positive value means that the visibility increased and a negative value means that the visibility decreased. For example, a gain of 4 means that the requester had 4 times more visibility after republication, while a gain of -3 means that the requester had 3 times less visibility. When a positive gain is observed, there is potentially a Streisand effect, since its republication potentially caused the requester to become more visible instead of less. Infinite gain means there is no visibility in the two months before the target date but visibility in the two months afterwards; negative infinity has an analogous interpretation.

6.3 Streisand Effect: Results

Google Trends did not have information for 58 of the 80 identified requesters, so we were only able to collect Google Trends information for 22 RTBF requesters, for which four of them are celebrities. Thus we see that many of the requesters are off Google Trends' radar screen, with few requests before or after republication. We were not able to collect tweets about two of the requesters due to the huge number of tweets about those requesters. On the other hand, some verified RTBF requesters were not mentioned at all. In total, we collected 23,731 tweets mentioning 44 out of 80 requesters, of which five of them are celebrities.

7 https://twitter.com/search-advanced

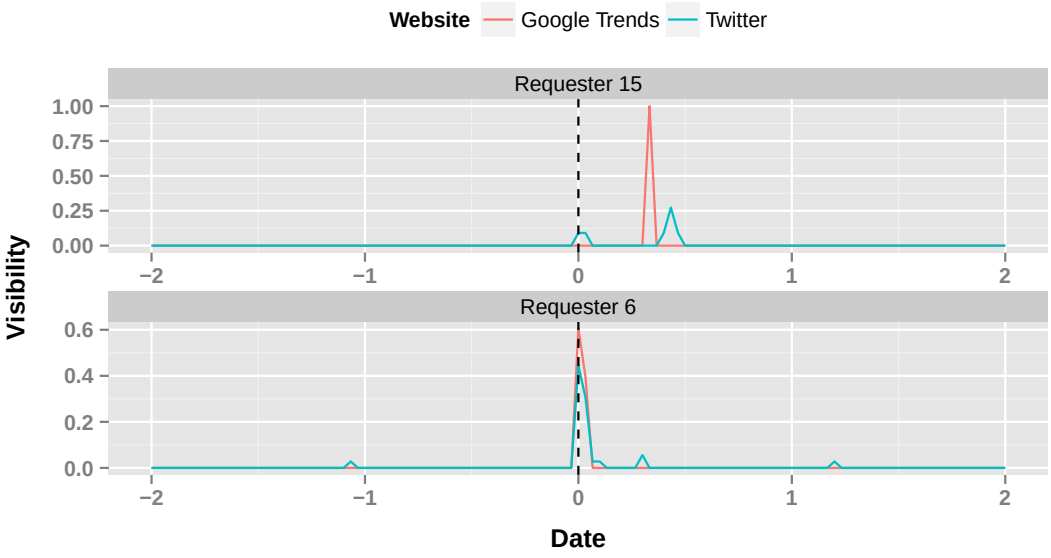


Fig. 6. Visibility as a function of time for two RTBF requesters. Dashed lines represent the republication dates of the articles.

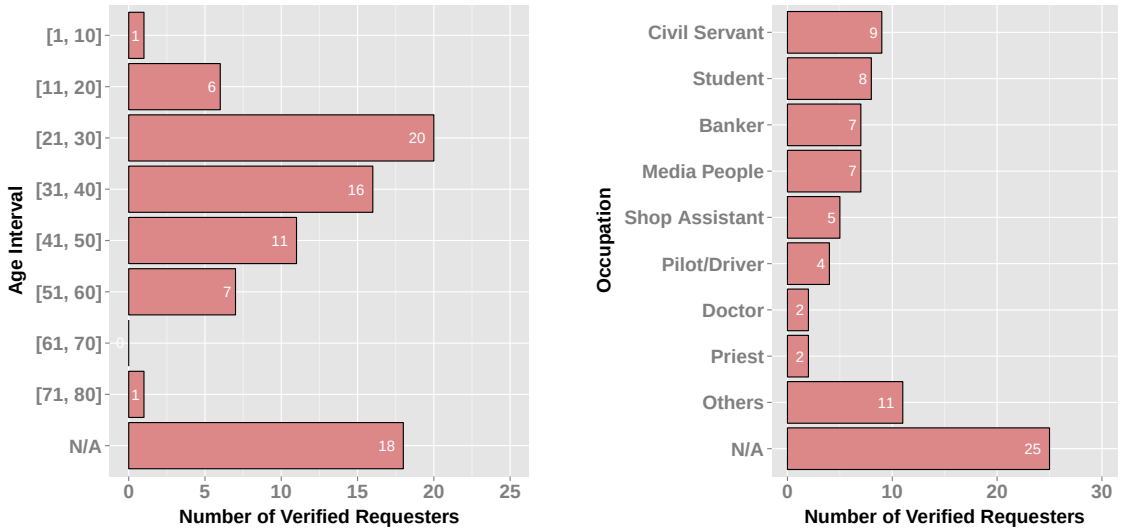
Table 2 tabulates the gain G for 22 requesters for Google Trends and 44 requesters for Twitter. We first observe that for both Google and Twitter, less than half of the requesters had a positive visibility gain. Thus, *the Streisand effect is not a general phenomenon for RTBF requesters with republished links*. We do see, however, that the average and maximum positive gains for Google Trends and especially Twitter are larger than the corresponding negatives.

Figure 6 shows the visibility as a function of time for two requesters with the significant positive gains in Google Trends and Twitter. The numbering on the x-axis corresponds to two months before and after republication, centered at zero. The vertical dashed lines are the republication dates of the associated links. We see that both Requester 15 and Requester 6 have almost no visibility two months before the republication date for both Google Trends and Twitter, but had peaks after the republication date for both Google Trends and Twitter. Thus, although the Streisand effect does not appear to be a general phenomenon for all requesters, some requesters may indeed suffer from it, although another plausible explanation is random noise. We also remark that if the news outlets published the requesters' names as well as the delisted URLs, the Streisand effect might be much more pronounced.

7 Demographic Study of RTBF Requesters

In this section we carry out a demographic analysis of the 80 verified RTBF requesters obtained in Section 5.1. We emphasize that this analysis only applies to *traditional media sites in the UK*, and the conclusions may not generalize to beyond the UK (and almost certainly will not to generalize to non-traditional media sites). For the demographic classifications for each requester, we simply just read the articles and do the classifications manually. For example, gender is determined by the first (given) name of the requester or by the gender indicated in personal pronouns [23]. (We were able to determine the gender for all 80 requesters.) The ages and professions of the requesters are also often provided in the articles themselves. In some cases we are not able to determine the age or profession of the requester directly from the article; for these cases, we label those values as N/A. We also put the names of the 80 requesters into google.com to check whether they are celebrities or not.

Strikingly, 87.5% (70 out of 80) of the requesters are male, which seems to imply that males are more inclined to make RTBF requests than females. However, this result is potentially biased by the fact that males may be mentioned more often than females in the mass media. To get a handle on this bias, we randomly select 1,000 BBC domestic articles spanning years 2000 – 2012 and relating to crime. From these articles, we extract the names in the articles, and determine the genders for the names, as described previously. In these 1,000



(a) Distribution of verified requesters by age

(b) Distribution of verified requesters by occupation

Fig. 7. Demographics of the 80 verified requesters

articles we find $1,689$ male names and 609 female names, so that approximately 73% of the names are male.

Because 87.5% of the requester names are male but only 73% of names in a typical article are male, it appears that men are more likely to make RTBF requests than females. We further validate this claim using hypothesis testing. Let x_R be the fraction of names that are male in RTBF articles (taken across the entire set of UK RTBF articles, not just those in our data set of 283 articles). Let x_N be the fraction of names that are male in arbitrary articles (taken across the entire set of UK articles). Owing to the aforementioned name analysis of the $1,000$ BBC articles, we can approximate $x_N = 0.73$. We want to show $x_R > x_N$ or equivalently $x_R > 0.73$. Our null hypothesis is

$$H_0 : \quad x_R = 0.73$$

That is, the fraction of names that are male in RTBF articles is equal to fraction of names that are male in normal articles. Consider sampling names from RTBF articles, and let X_i be equal to one if the ith sampled name is male; zero if female. Each X_i is independent and identically distributed (i.i.d.) Bernoulli random variable under H_0, having mean 0.73. Let $Y = X_1 + X_2 + \cdots + X_{80}$, that is, the total number of names that are male from a sample of 80 RTBF requesters. Since the X_i's are i.i.d. Bernoulli, it follows that Y is a binomial distribution with parameters 80 and 0.73. We can now test the hypothesis. We observed 70 males in our sample. We can calculate:

$$P(Y \geqslant 70 \mid H_0) = 0.0014 = p\text{-value}$$

Thus the probability of observing 70 or more males under the hypothesis H_0 is very small, and much less than 0.05 (typically used for statistical significance). Therefore we can reject the null hypothesis, accept the alternative $x_R > 0.73$, and claim that men are more likely to make RTBF requests (for media links) than females with statistical significance.

Figure 7 summarizes the demographic results for age and occupation. For age shown in Figure 7(a), there is great diversity, but the majority of the requesters are between ages 20 and 40. For occupation, we categorize the requesters into ten types of occupations, which are doctor, priest, pilot/driver, shop assistant, media people, banker, student, civil servant, others, and N/A. In Figure 7(b), the ten types of occupations are ordered by decreasing number of verified requesters. The most common occupations in the data set are "Civil Servant" (including police officers and mail persons), "Student," "Banker," and "Media People." We also found that 14 of the 80 requesters are celebrities. The large majority of the requesters in the data set are more ordinary people.

8 Related Work

Before the RTBF became law, legal theorist and economist Viktor Mayer-Schönberger wrote about perfect remembering in the digital age, and why we must reintroduce our capacity to forget [19]. Since the RTBF became law, it has received extensive media coverage on a global scale as well as some scholarly attention. Legal theorists have carried out significant

work on the RTBF, with a focus on the tradeoff of the privacy benefits versus potential infringements on freedom of speech: see Bolton [8], Ambrose [1], Ambrose and Ausloos [2], Weber [29], and Koops [6]. From an economic perspective, Kim *et al.* [16] develop game-theoretic models for the RTBF, and argue that the global expansion of the RTBF should not be taken as a threat to the right of free speech and access to information nor can it be justified as an effort to strengthen privacy rights. Rather, it should be understood by analyzing the optimally balanced level of protecting both privacy rights and freedom of speech as represented by the socially optimal level of link removals.

Data-driven studies provide quantitative insights that can inform the debates surrounding the RTBF. However, to date, little work has been done in this direction, particularly in the research community. On its transparency page, Google itself provides some data analysis results, including data on requests accepted and rejected from the various EU countries, and data on the most effected by the RTBF [14]. Tippman and Pamis create visualizations of data accidentally revealed by Google, and quantify the percentage of the requests that are for basic privacy concerns such as addresses and sexual affiliation [27].

To the best of our knowledge, this is the first data-driven study for RTBF based on actual delisted content. It is the first paper to quantitatively study the delisted content, the demographics of requesters, and the potential Streisand effects the RTBF may engender. It is also the first paper that raises the possibility of data-driven cyber attacks against the RTBF; the paper analyzes the potential damage these attacks can cause, and the effort required to carry out these attacks. We believe the results and observations in this paper can inform lawmakers as they refine the RTBF laws in the future, and in particular, inform the current debate on whether the RTBF law should be extended beyond the European Google sites to all of Google's sites.

9 Conclusion and Recommendations

In this paper we undertook a data-driven study of the RTBF. We analyze the content of 283 known delisted links, devise data-driven attacks to uncover previously-unknown delisted links, and use Twitter and Google Trends data to investigate whether rediscovered RTBF requesters can suffer from the Streisand effect. One of the values of scientific work is to be able to predict results for large-scale systems with relatively small samples. Although our data set is only moderate in size with 283 articles, it is large enough to allow us to gain meaningful insights, identifying factors and variables that impact the results and limitations of the RTBF ruling, perhaps the most important privacy law of the decade.

9.1 Opinions and Recommendations

We end this paper with a few opinions and recommendations based on the results and observations from this paper. After having studied RTBF and its consequences from a data perspective, the authors feel that RTBF has been largely working and responding to legitimate privacy concerns of many Europeans. We feel that Google's process for determining which links should be delisted seems fair and reasonable. We feel that Google is being fairly transparent about how it processes RTBF requests [13]. Other academics have called more transparency [12]. However, by being more specific about how the delisting decisions are made, it may become easier for the attacker to rediscover delisted URLs and the corresponding requesters.

We strongly recommend that Google desist from notifying the webmasters about their delisted content. As we have seen in this paper, some of these media companies are republishing the links, essentially acting as transparency activists. Furthermore, we showed in this paper that many of the requesters can be identified from the republished delisted links.

Although the RTBF has been largely working to date, we also feel that the attacks described in this paper put the efficacy of the law into question for online-media content. We have shown that a third-party (such as a transparency activist or a private investigator) can rediscover RTBF delisted URLs for traditional media sites, and we have argued that with sufficient resources, the attacker can potentially rediscover $30\% - 40\%$ of RTBF delisted URLs in traditional media sites. A transparency activist could then republish these URLs along with the requesters' names, thereby generating a Streisand effect. Moreover, we do not see any effective defenses to these attacks, except for delisting the articles no matter what the query – a defense many people would consider to be a strong form of censorship [18].

We also feel that lawmakers should exercise greater caution when creating new laws addressing online privacy. As shown in this paper, and an early paper on the Children's Online Privacy Protection Act (COPPA) [11], due to big-data inference, online privacy laws can potentially have the unintended consequence of reducing individuals' privacy rather than protecting it.

Acknowledgements

We thank Kami Vaniea and the anonymous reviewers for their valuable feedback. This work was supported in part by the NSF under grant CNS-1318659. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of any of the sponsors.

References

[1] M. L. Ambrose. Speaking of forgetting: Analysis of possible non-EU responses to the right to be forgotten and speech exception. *Telecommunications Policy*, 38(8–9):800–811, Sept. 2014.

[2] M. L. Ambrose and J. Ausloos. The right to be forgotten across the pond. *Journal of Information Policy*, 3, 2013.

[3] J. Ausloos. The "Right to be Forgotten" – Worth remembering? *Computer Law & Security Review*, 28(2):143–152, 2012.

[4] BBC. List of BBC web pages which have been removed from Google's search results. http://www.bbc.co.uk/blogs/internet/entries/1d765aa8-600b-4f32-b110-d02fbf7fd379, 25 June, 2015.

[5] BBC. BBC forgotten list "sets precedent". http://www.bbc.com/news/technology-33287758, 26 June, 2015.

[6] Bert-Jaap Koops. Forgetting footprints, shunning shadows: A critical analysis of the "Right to be Forgotten" in big data practice. *SCRIPTed*, 8(3):229–256, Dec. 2011.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[8] R. L. Bolton. The right to be forgotten: Forced amnesia in a technological age. *John Marshall Journal of Information Technology & Privacy Law*, 31(2):133–144, 2015.

[9] M. Cacciottolo. The Streisand effect: When censorship backfires. http://www.bbc.com/news/uk-18458567, 2012.

[10] CNIL. CNIL orders Google to apply delisting on all domain names of the search engine. http://www.cnil.fr/english/news-and-events/news/article/cnil-orders-google-to-apply-delisting-on-all-domain-names-of-the-search-engine/, 2015.

[11] R. Dey, Y. Ding, and K. W. Ross. Profiling high-school students with facebook: how online privacy laws can actually increase minors' risk. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 405–416. ACM, 2013.

[12] E. P. Goodman. Open letter to Google from 80 Internet scholars: Release RTBF compliance data. https://medium.com/@ellgood/open-letter-to-google-from-80-internet-scholars-release-rtbf-compliance-data-cbfc6d59f1bd#.at81h9i60, 2015.

[13] Google FAQ. Google frequently asked questions: European privacy in search. https://www.google.com/transparencyreport/removals/europeprivacy/faq, 2016.

[14] Google Transparency Report. Google transparency report: European privacy requests for search removals. https://www.google.com/transparencyreport/removals/europeprivacy, 2016.

[15] Hidden From Google. http://hiddenfromgoogle.com/, 2014.

[16] B.-C. Kim and J. Y. Kim. The Economics of the right to be forgotten. *NET Institute Working Paper*, 2015.

[17] E. Lee. The right to be forgotten v. free speech. *Free Speech (August 26, 2015). Chicago-Kent College of Law Research Paper Forthcoming*, 2015.

[18] H. J. Lee, J. H. Yun, H. S. Yoon, and K. H. Lee. The right to be forgotten: Standard on deleting the exposed personal information on the Internet. In J. J. Park, I. Stojmenovic, H. Y. Jeong, and G. Yi, editors, *Computer science and its applications*, pages 883–889. Springer, 2015.

[19] V. Mayer-Schönberger. *Delete: the virtue of forgetting in the digital age*. Princeton University Press, 2011.

[20] M. Mondal, B. Viswanath, A. Clement, P. Druschel, K. P. Gummadi, A. Mislove, and A. Post. Defending against large-scale crawls in online social networks. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pages 325–336. ACM, 2012.

[21] A. L. Newman. What the "Right to be Forgotten" means for privacy in a digital age. *Science*, 347(6221):507–508, 2015.

[22] M. L. Rustad and S. Kulevska. Reconceptualizing the right to be forgotten to enable transatlantic data flow. *Harvard Journal of Law and Technology*, 28:349, 2015.

[23] C. Tang, K. Ross, N. Saxena, and R. Chen. What's in a name: a study of names, gender inference, and gender behavior in facebook. In *Database Systems for Adanced Applications*, pages 344–356. Springer, 2011.

[24] The Guardian. Google to extend "Right to be Forgotten" to all its domains accessed in EU. http://www.theguardian.com/technology/2016/feb/11/google-extend-right-to-be-forgotten-googlecom?CMP=twt_a-technology_b-gdntech, 2016.

[25] The New Yorker. The solace of oblivion. http://www.newyorker.com/magazine/2014/09/29/solace-oblivion, 2014.

[26] The New Yorker. Google will further block some European search results. http://www.nytimes.com/2016/02/12/technology/google-will-further-block-some-european-search-results.html?_r=1, 2016.

[27] S. Tippmann and S. Pamiés. Google's data on the right to be forgotten. http://sytpp.github.io/rtbf/index.html, 2015.

[28] A. Tsesis. The right to erasure: Privacy, data brokers, and the indefinite retention of data. *Wake Forest Law Review*, 433(49), 2014.

[29] R. H. Weber. The right to be forgotten: More than a Pandora's box? *Journal of Intellectual Property, Information Technology and E-commerce Law*, 2:120–130, 2011.