HALMSTAD UNIVERSITY

Master's in Network Forensics

MASTER THESIS

An analysis of the Privacy Policy of Browser Extensions

Susan Sarah Zachariah

School of Information Technology

Abstract

Technological advancement has transformed our lives by bringing unparalleled convenience and efficiency. Data, particularly consumer data, essential for influencing businesses and developing personalized experiences, is at the heart of this transition.

Companies may improve consumer satisfaction and loyalty by using data analysis to customize their products and services. However, the collection and utilization of consumer data raise privacy concerns. Protecting customers' personal information is essential to maintaining trust, respecting individual autonomy, and preventing unauthorized access or misuse.

Along with the protection of data, transparency is also another essential factor. When companies or organizations deal with users' data, they are liable to inform these users of anything and everything that happens with their data.

Our study focuses on the online privacy policies of Google Chrome browser extensions. We have tried to find the extensions that comply with the data protection guidelines and if all Google Chrome browser extensions are transparent enough to mention the details as per guidelines. Utilizing the power of Natural Language Processing (NLP) techniques, we have employed advanced methodologies to extract insights from these policies.

Keywords: Privacy policy, Browser extension, Data privacy and management

Acknowledgments

I want to express my heartfelt thanks to my supervisor, Pablo-Picazo Sanchez, who was instrumental in my selecting this topic. His constant guidance and supervision is the reason I could complete this project.

I sincerely thank esteemed examiners Mark Dougherty and Eric Jarpe. Their timely review and feedback were insightful and helped me in the research.

Acronyms used

- NGOs: Non-government organizations
- GDPR: General Data Protection Regulation
- COPPA: Children's Online Privacy Protection Act
- CCPA: California Consumer Privacy Act
- HIPAA: Health Insurance Portability and Accountability Act
- APP: Australia's Privacy Principle
- NLP: Natural Language Processing
- NLTK: Natural Language Tool Kit
- PII: Personally Identifiable Information
- POS: Part of Speech
- NER: Name Entity Recognition
- DBDP: Data Broker and Data Processor
- LDA: Latent Dirichlet Allocation

Contents

Figures

Tables

1 Introduction

Technological developments lead to innovations in both business and everyday life. With technological advancement, especially over the last decade, there has been an explosion of information about people, their behaviour and their contexts. This phenomenon is referred to as learning analytic data, as the data collected are analyzed and interpreted to personalize the experience.

The production and the consumption of learning analytic data have been a significant factor in transforming organizational practices, opening new pathways for data collection, analysis and reporting that previously seemed impossible. As a result, social actors can now use more data in a targeted way and in real-time. Social actors may include various entities, including individuals, groups, organizations and institutions like government agencies, private companies, non-government organizations (NGOs), community groups and even influencers. With new emerging technologies and methods, a vast amount of data, including personal and confidential information of the users, is now easily accessible to these social actors.

Data are a vital resource that drives decision-making processes, supports the functionality of various services and technologies and creates innovation. Multiple business institutions and governments use them to study consumer behavior, enhance user experiences, streamline operations and make informed strategic decisions. The advent of machine learning and artificial intelligence has helped automate processes and personalize user services by using algorithms that predict trends and user patterns by expanding the potential of data, making data more and more valuable.

The modern web, or in other words, the Internet, serves as a host for a myriad of online services. Web browsers provide an interface between the user and the online content through the Internet. They serve as a gateway to the immense amount of information available on the Internet by playing a crucial role in shaping the user's digital experience. They request users for their personal information; for example, most popular websites require users to register by giving their details to create an account, including their email address, name and date of birth. Another example is that many social media platforms witness users sharing vast amounts of personal information, like posting individual or family photos online, revealing information like places checked in and

products purchased or used. In short, users willingly divulge much of their personal information online.

The features of the web browsers can be enhanced using browser extensions. The browser extensions can interact with web pages, change the browser's appearance and provide new functionality through the user interface. Because of the privileged positions of browser extensions inside a user's browser, they have access to the content and personal data like form entries, passwords, history, web cache and cookies of users [22, 56]. This personal information is gathered, matched, transferred and profiled as a part of the routine engagement by the websites, which sometimes fail to handle the personal data of the users in a discreet manner, potentially exposing users to privacy risks.

The ever-evolving landscape of the modern web presents a complex relationship between user engagement, utilization of personal data by online services and the potential risks introduced by websites (browser extensions, in our case). As users navigate the digital realm, awareness of these dynamics becomes paramount, urging those involved to prioritize the responsible use and management of personal information in the ever-expanding online ecosystem.

Governments and regulatory bodies worldwide now recognize the importance of privacy protection and are enacting or updating regulations to address these concerns. Accordingly, there have been recent changes in the regulations, especially the principles in Article 5, section 2 of the General Data Protection Act (GDPR), designed to restrict how data are managed [47]. In the context of data protection, the countries in the European Union follow GDPR guidelines. In the United States, although there is no comprehensive national privacy law, there are data protection laws that are sector-specific, like the Children's Online Privacy Protection Act (COPPA), California Consumer Privacy Act (CCPA), Health Insurance Portability and Accountability Act (HIPAA) [42]. Similarly, different countries have their guidelines when it comes to the protection of data, like Australia's Privacy Principles (APP) in Australia, the Information Technology Act in India, the Electronic Communications and Transactions Act in South Africa, the Personal Data Protection Act in Singapore [48]. However, most of these guidelines are influenced by GDPR [27].

According to GDPR guidelines, companies must obtain explicit consent from individ-

uals before collecting and processing their data [45]. Privacy policies are the primary channel through which companies inform users about their data collection and sharing practices [29].

As per Google Chrome's Web Store developer agreement [24], the developer has to ensure that users are informed about data collection if their extension is doing that and violation of the agreement or Google Chrome Web Store Program Policies may lead to suspension or termination of the rights to publish on the Web Store as per clause 8.3. Therefore, to comply with the directions of the developer agreement, websites typically post privacy policies based on the **notice-and-choice principle**; that is, they show users the terms of the privacy policy and the users can accept or reject the terms mentioned in the privacy policy [57]. However, in practical situations, users do not read the complete privacy policies on websites or have difficulties understanding them [57].

Hence, data-driven evolution creates ethical and privacy considerations. Poor privacy decisions by corporates and individuals may lead to undesirable consequences, such as selling personal data or information to unknown third parties or unintentionally collecting personal data for use in newsletters, analytic, personalization or even phishing attempts [47].

Recently, authors analyzed 178,893 extensions from the Chrome Web Store between September 2016 and March 2018 and 2,790 extensions from the Opera browser [14]. The analysis showed that 3,868 (2.13%) extensions potentially leaked privacy-sensitive information. Also, among the Chrome browser extensions confirmed to leak privacy-sensitive information, the top ten extensions alone have over 60 million users combined.

In this MSc project, we analyze the privacy policies of Google Chrome browser extensions and answer the following questions:

RQ1 Do all browser extensions in the Google Chrome web store inform users about their data collection and management practices through a separate privacy policy?

RQ2 How many Google Chrome web store browser extensions use the same privacy policies?

2 Background

In this section, we introduce terms like browser extension and privacy policy that we will use in the report.

2.1 Browser Extensions

A web browser is a software application used to access and view web pages on the Internet. A few examples of web browsers include Google Chrome, Mozilla Firefox, Apple Safari, Microsoft Edge and Internet Explorer [43]. The user's browsing experience while accessing web pages can be enhanced or customized as per user preference using browser extensions.

A browser extension is a piece of software installed as an add-on in the web browser to offer additional features like ad blockers, language translators, sound boosters, video editors and password managers. It can interact with web pages, change the browser's appearance and provide new functionality through its user interface. Because of the privileged position of browser extension inside a user's browser, they might have access to the content and functionality not available to web pages, such as the ability to conduct and read cross-origin requests and access a browser's history and cookies [10]. The privileged access of browser extensions in a web browser can cause privacy concerns, even in private browsing mode [56].

Browser extensions are found in web browser's official extension stores or marketplaces. For instance, the Chrome Web Store is the go-to place for extensions for Google Chrome, while Firefox Add-ons is where we can find extensions for Mozilla Firefox. Before installing a browser extension, it is always advisable to read its reviews, ratings, attached policies and terms & conditions. A browser extension's privacy policy specifies how the developer collects, stores and uses personal information gathered through their extension. This data could include browsing history, IP address, location information and cookies.

The Browser extensions extend the functionality of web browsers that interact with individuals, while privacy policies govern how the extension developer acts or is supposed to function.

There are many browser extensions available that provide benefits to users like [19, 20]:

a. **Customization:** Browser extensions allow users to customize their browser experience according to their preference, adding features and functions not built into the browser.

b. **Enhanced productivity:** Browser extensions are available that can increase productivity by streamlining tasks, automating processes and adding functionality that improves workflow.

c. **Accessibility:** Browser extensions can make web content more accessible for users with disabilities by providing features such as text-to-speech, screen readers and other assistive technologies.

d. **Security:** Browser extensions can enhance security by blocking ads and pop-ups, scanning for malware, protecting user privacy by blocking and tracking cookies and employing other data collection methods.

e. **Entertainment:** Some browser extensions provide entertainment value by adding games, social media features or other fun elements to the browsing experience.

While browser extensions can enhance the user's browsing experience by providing additional features, they can pose privacy and security threats [26, 31, 53]. For example:

i. **Malicious extensions:** Malicious browser extensions may download malware, spyware or Trojan horse virus in the system that can harm the computer and steal personal information.

ii. **Data collection:** Browser extensions may collect browsing history, including search queries, websites visited, sensitive information, and personal data and sell them to third parties.

iii. **Security vulnerabilities:** Browser extensions may have security vulnerabilities that hackers can exploit and gain access to the computer to steal personal information.

iv. **Compatibility issues:** Browser extensions may not be compatible with the browser or existing extensions used in the particular system, causing stability issues or other technical problems.

v. **Performance issues:** Multiple browser extensions may consume many system resources, causing web browsers to slow down or crash [10].

2.2 Privacy Policies

A privacy policy is a legal document that outlines "what" and "how" a company or organization collects, uses, shares and protects the personal information of its customers or users.

According to GDPR guidelines, individuals have the right to know what personal data companies collect about them, how they use it and with whom they share it. Individuals also have the right to access, modify or delete their data, object to its processing or request its transfer to another company. Services or businesses that collect this information must maintain the confidentiality of personal data collected for transactions within EU member states [32] [Information Commissioner's Office, licensed under the Open Government Licence v3.0 For Organisations – A guide to individual rights, updated as of 19 May 2023].

Browser extensions display their privacy policy on the web page, but the placement of the policy document or link varies from site to site and is not a fixed location. It is usually available either on the extension's download page, the developer's website, the settings menu or the options menu [25]. There are also many browser extensions without a published privacy policy. Some browser extensions have privacy policies uploaded in their local language and are unavailable in English. Companies usually provide a link to their privacy policy during the account registration process when users sign up for newsletters or subscribe to any offers or advertisements. In addition to being available on a company's website, privacy policies are also included in additional documentation, such as terms of service and end-user license agreements or as part of their "terms of service" in the case of mobile apps.

2.2.1 Importance of Privacy Policy

We can summarize the importance of privacy policies below [37]:

i. **Transparency:** A privacy policy promotes transparency between a company and its customers or website visitors. It also helps build users' trust in the company.

ii. **Protection:** A privacy policy outlines the measures a company takes to protect the personal information of its customers or website visitors. It can include techniques like encryption, firewalls, and other security measures that help prevent unauthorized access or data breaches.

iii. **Accountability:** A privacy policy provides a clear framework on how a company handles personal information and helps them hold accountable if there is a breach or misuse of data.

iv. **Compliance:** Many countries and regions have laws and regulations requiring companies to have a privacy policy, such as GDPR in the European Union. Failure to comply with these regulations can result in legal penalties. Having a privacy policy ensures that a company is compliant with these laws.

v. **User rights:** A privacy policy outlines the user's rights over their personal information, such as the right to access, modify and delete their data. It helps to empower users and give them control over their data.

3 Related Work

As this is an emerging field, limited research and studies have been done in the area related to privacy policies of browser extensions. In this paper, we present our analysis of privacy policies of Google Chrome Web Store browser extensions. In this section, we have given a gist of a few studies about privacy policy analysis, studies on browser extensions and later, we have tried to merge both and present our results.

3.1 Privacy Policy

Existing work on privacy policies shows analysis from varied angles to study its importance, different perspectives and implications on data collection and management practices. The studies have been done on both policies that were available online and offline.

Researchers have conducted various studies and analyses of privacy policies based on different categories, including but not limited to privacy policies of Android apps [55, 50], iOS apps [21], smart home devices [39] and browser extensions [14, 12].

In a paper by Juniper Lovato, Philip Mueller, Parisa Suchdev and Peter S. Dodds [38], the authors have examined the privacy policies of more than two decades. They developed a dictionary of PII (Personally Identifiable Information) related terms from their extensive data set, which was textual privacy policies obtained from Amos et al. that included over one million privacy policy snapshots from over 100,000 websites from 1997 to 2019. The analysis results that were obtained at the word level highlighted the stability of privacy policy PII data type terms over time; at the topic level, the results highlighted the complexity of the privacy policies, which was measured by comparing the compression factor with the previous years, showing that the complexity of privacy policies has decreased over time and at network level the results of the analysis were showcasing the sensitivity and risk in the privacy policies by measuring the density of the word co-occurrence to understand privacy policies' stability, complexity and sensitivity over time.

A considerable number of studies have been conducted in terms of their readability, complexity and comprehension [1, 9, 33, 40]. Earlier research has been done to identify

proper purpose-centric statements, vagueness in the text of policies and contradictions in the policy and actual practice [52, 3, 6, 55, 2, 17, 15]. Studies have also been performed to know the options provided to the user based on the "Opt-out choices" and "notice and consent" facility provided to the user [44, 49]. Another area on which privacy policy studies have been conducted is compliance with the guidelines [8, 9, 15].

3.2 Browser Extension

The present Chrome extension system stems from the design proposed by Barth et al. [5]. The proposed design was aimed more at the developers and could have been easily exploited by malicious website operators. In recent years, much research has focused on security and permission models in light of the possible vulnerabilities [4, 5, 13, 28]. Research has been conducted [35] to dynamically analyze Chrome browser extensions and detect malicious behaviour in browser extensions. In their paper, the authors confirmed identifying 130 malicious and 4,712 suspicious extensions with up to 5.5 million browser installations.

In a related paper [36], the authors stated that the browser extensions obtain user permission to collect, store and process the user's personal data during installation. However, companies divulge very limited information regarding the data collection and management practices and the potential risks to the users. According to [14], 2.13% of extensions may have exposed privacy-sensitive information and the top 10 most popular Chrome extensions with privacy issues have over 60 million users combined. The study gives insight into the vast amount of information leak that occurs through these browser extensions alone. In [56], the authors point out that even in private browsing mode, there are privacy breaches that are caused by some browser extensions.

This study will analyze the privacy policy of Google Chrome Browser Extensions.

3.3 Automatic Analysis of Privacy Policies

Earlier research has suggested that most times, the privacy policies are written in a "legal language", making it difficult for average users to understand them [57, 33, 46, 50]. The relevant information can be extracted from the text of these privacy policies using Natural Language Processing (NLP) [59] and deep-learning techniques [30, 51]. These

techniques extract the relevant data and present the information in more accessible ways to the users [7, 18, 30, 41, 54]. Some studies have used several pre-defined patterns to analyze privacy policies. In the paper by Costante et al. [16], they developed a technique for automatically analyzing privacy policies. It also displays the information collected by the websites based on the information extraction techniques used. Brodie et al. [11], mentioned that they created a method to convert written privacy policies into machine readable form by creating a set of grammars and a specific set of rule elements using NLP. In paper [55], the authors claim to have developed a system called PPChecker that would automatically discover the inconsistencies in the Privacy policies. Zimmeck et al. [58], proposed a software architecture called Privee that would be able to determine whether a privacy policy contains statements related to information collection. On the other hand, Slavin et al. [50], proposed a semi-automatic method to find the information retained in the app's byte code but failed to mention it in its privacy policy. Such privacy policies would fall under the category of Incomplete Privacy Policies.

4 Research Methodology & Analysis

This chapter introduces the methods followed for obtaining the data set and details about the analysis in the study. Given the popularity of Chrome, we restricted our analysis to the browser extensions stored in the Web store, the public repository maintained by Google from which users can install browser extensions.

In the following sections, we have explained how we obtained the privacy policy of all the browser extensions and leveraged this data to extract results and meaning from the data set. During various stages, we have used open-source tools wherever possible.

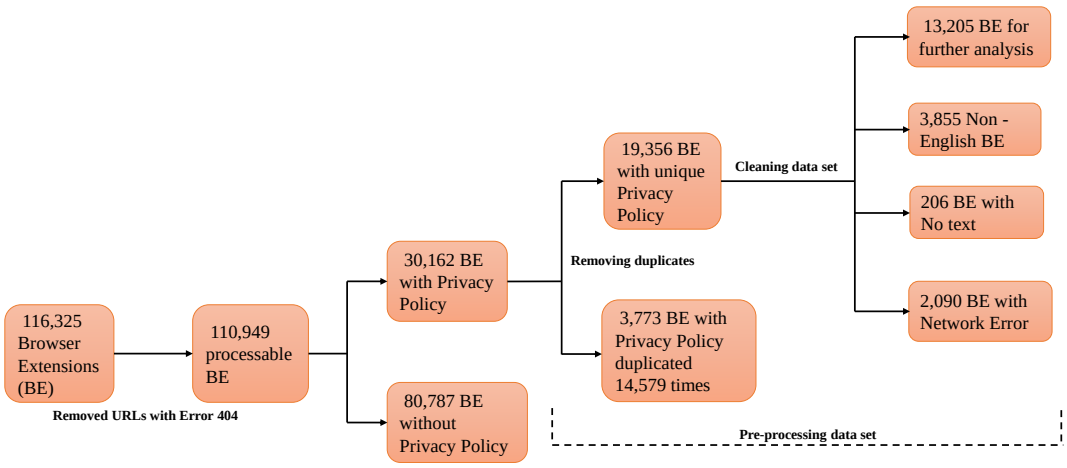


Figure 1: Overview of Data set

The Google Chrome Web store has more than 100,000 different browser extensions. These extensions are being added, deleted and updated continuously. A few extensions have also become redundant and are not in use. Figure 1 depicts the overview of the data set (explained in detail in the following section) that we have used in this study.

4.1 Dataset

We crawled the Google Chrome Web Store and obtained links for 116,325 browser extensions as of February 2023. We used the Python script for the purpose.

Of all the browser extension links obtained, few extensions were redundant or had Network error, preventing us from further processing those links. We discarded 5,376 (4.60%) such browser extension links and proceeded with the remaining 110,949 browser extensions (as in Figure 1), which forms our base data set, in May 2023.

On repeating the process in October 2023, we discarded a total of 21,485 (18.50%) browser extension links, as they had become redundant or had Network error and proceeded with the remaining 94,840 browser extensions, which forms our second base data set, in October 2023.

4.2 Collecting information of dataset

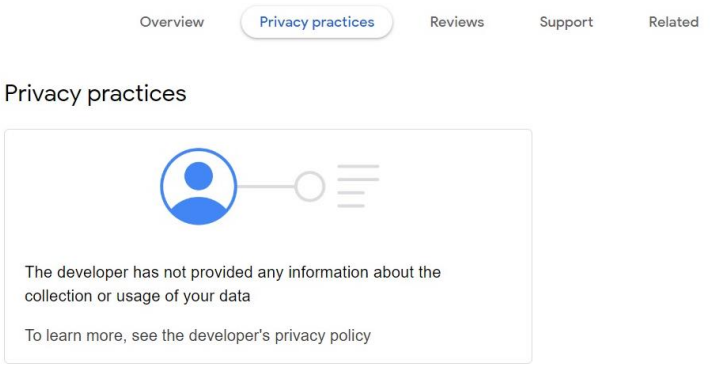
For every browser extension, we collected information such as name, updated date, size, contact information, number of users and privacy policy URL, wherever available, first in May 2023 and again in October 2023.

4.2.1 Collecting data from web page

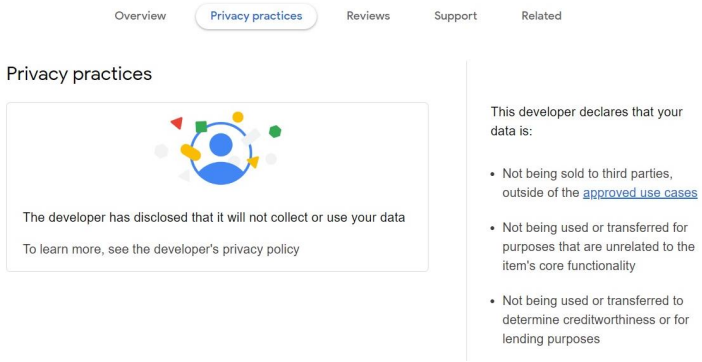
The list of browser extensions to be scraped for further fetching their details is maintained as an HTML file. Using the Beautiful Soup library, we extracted attributes from the web page based on the tag and the required text. For example, for extracting "name" information, the code looks for <title> tag within <head> section of the particular HTML document. It extracts content from the <title> tag and assigns it to the variable "name". Similarly, to extract privacy policy URLs, the code looks for an <a> tag with the text "Privacy Policy" and extracts the value of the particular href attribute. The corresponding URL is then assigned to the variable "Privacy Policy".

Similarly, we collected all the data available on the web page of the browser extensions.

Based on the information collected, we found that the majority of the browser extensions from the Web Store show a declaration stating that the developer has not provided any



(a)



(b)

Figure 2: No Privacy Policy Declaration

information regarding collection and usage practices to Google Chrome (as in Figure 2a) or a declaration that the developer will not be collecting any data (as in Figure 2b).

Even though there are specific guidelines stating the requirement for providing details about data collection, management and sharing practices, the developers and website owners find loopholes in the system to avoid giving all users accurate information.

4.3 Pre-processing the data set

After collecting browser extensions with separate privacy policy URLs, our next major step is cleaning and arranging the data for further processing.

4.3.1 Removing Identical URLs

We segregated browser extensions having separate privacy policy links. Using Excel, we removed all the repetitive links based on the updated date, i.e., when more than one browser extension has the same privacy policy link, we eliminated the older ones and maintained the one with the latest updated date.

After removing the duplicated or identical privacy policy URLs and the corresponding browser extensions, we have obtained our set of unique individual privacy policies, which will be processed further.

We feel that multiple browser extensions have the same privacy policy because, instead of having individual privacy policies for all the browser extensions developed by the developer, it would be much more convenient for the developer to maintain a standard privacy policy and the link for the particular privacy policy is provided to all the extension he develops. Also, if the privacy policy is a standard comprehensive policy, including all aspects of data collection, management and sharing practices, it would be convenient for developers to use the same privacy policy, especially if the number of extensions developed are in large numbers.

4.3.2 Cleaning the data set

After obtaining the data set containing unique privacy policy URLs of browser extensions, we cleaned the data set before we could finally extract text from the privacy policies for our analysis.

For this purpose, we have used Beautiful Soup and langdetect libraries.

We have limited our studies to English as evaluating privacy policies in other languages would require additional language proficiency. For this purpose, i.e., to exclude non-English privacy policies from our crawls, we ran a language detection crawl that segregated non-English privacy policies in a separate file.

Also, certain privacy policy URLs had no text in the link given as privacy policy. All such privacy policies with similar exceptions were categorized as errors for this study and were eliminated from further processing.

5 Results & Discussion

In this chapter, we analyze and present the study results about privacy policies of browser extensions. We also offer our analysis, which will help us to answer our research questions.

Based on analysis details of Section 4.2, we found that out of the May 2023 base data set of 110,949 browser extensions, 27% (30,162) had links to an external privacy policy, i.e., the privacy policy is not in the Web Store but allocated in another Internet address. For the data set of October 2023, the figures have mostly stayed the same. It is observed that 28.50% (27,046) had links to a separate privacy policy, as depicted in Figure 3.

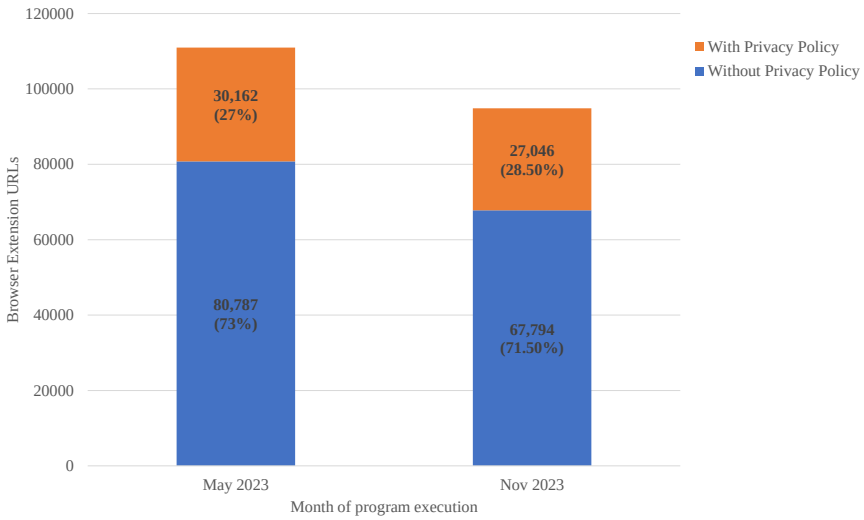


Figure 3: URLs with and without Privacy Policy

From Figure 3, we can confirm that all browser extensions in the Google Chrome Web Store do not inform users about their data collection and management practices through a separate privacy policy. Less than 30% of the browser extensions adhere to this practice and have a separate privacy policy even though specific guidelines like GDPR and CCPA emphasize the importance of privacy policies and require website owners to be transparent about the information they collect from the users and obtain user consent. The result of this study indicates a notable gap in adherence to the basic concept of

transparency regarding data collection and management practices. It is crucial that users are informed about how their data is handled and the lower percentage suggests that a substantial portion of extensions may lack explicit communication on these matters.

According to an article published in February 2019 [34], the study conducted by a security firm Duo reported that 85% of extensions and apps in the Google Chrome Web Store do not have a listed privacy policy [34, 23]. With the onset of COVID-19 around November 2019, the usage and dependency on online platforms increased considerably, followed by strict adherence to online regulations. It could be one of the factors contributing to more developers uploading their privacy policies, raising the percentage of presence of privacy policies from 15% before February 2019 to 27% in May 2023.

However, [34] also mentions that the absence of a privacy policy does not necessarily mean the extension is malicious. Still, it makes it difficult for users to understand what data is being collected and how the collected data is being managed.

One interesting observation during our analysis was that 4,283 Chrome browser extension URLs were updated at least once in six months, i.e., between May 2023 and October 2023. However, they still do not have a separate privacy policy.

Of the 30,162 browser extension URLs with a privacy policy (as of May 2023), 3,773 were duplicated and used for 14,579 browser extensions, leaving only 19,356 unique privacy policy URLs, which addresses our RQ2. The analysis indicates that multiple browser extensions share the same privacy policy, possibly maintained by the same developer for convenience. Even though maintaining a uniform and standard privacy policy might be convenient for developers, it raises concerns about the data collection and management practices of browser extensions. It also raises concerns about whether these shared policies accurately reflect the unique data practices of each extension.

The repetition of privacy policies suggests a common approach among the developers, that they are potentially using a standard comprehensive policy for multiple extensions. It also emphasizes the need for closer examination of shared privacy policies to ensure they adequately address each extension's specific data collection and management practices. It also indicates a potential challenge in providing users with precise and tailored information about how each extension handles their data.

Figure 4 shows the graph with the number of times different browser extensions use the same privacy policy. (Due to size restrictions, we have been able to include only a few privacy policies.)

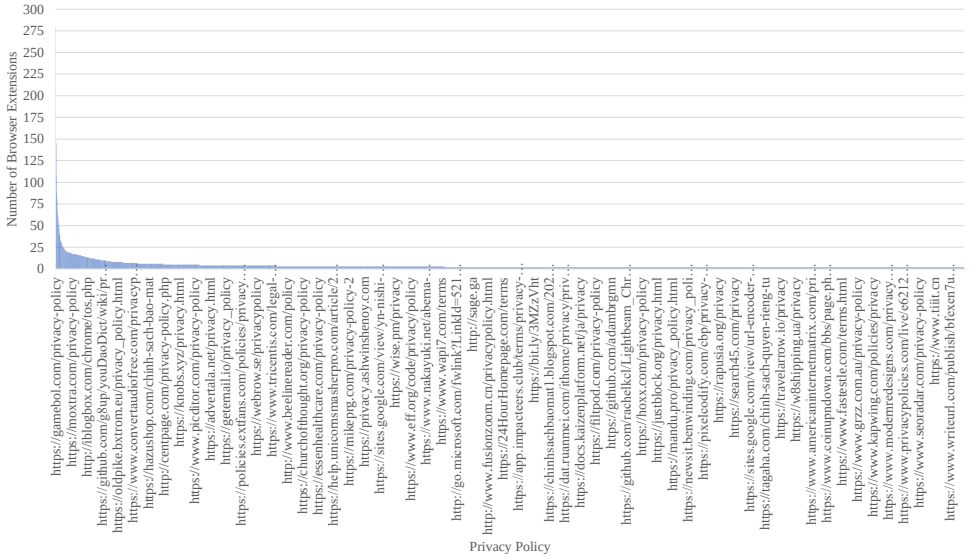


Figure 4: Browser Extensions with same Privacy Policy

The highest number of times the same privacy policy has been repeated was 279 times for various browser extensions. In Table 1, we include the top 5 of the most used Privacy Policy URLs and the number of browser extensions that used the same privacy policy.

Privacy Policy URL	**Extensions**
https://gamebol.com/privacy-policy	279
https://724fun.com/privacy-policy-2	147
https://pickergame.com/page/privacy-policy	144
https://www.loupdb.com/privacy-policy	107
http://foxexpress.com.vn/privacy-policy.html	88

Table 1: Duplicated Privacy Policies

Hence, we can conclude the result of RQ2 is that 3,773 privacy policies were used for

14,579 browser extensions in May 2023.

To further process the 19,356 unique privacy policies, we segregated the 13,205 browser extension privacy policy for analyzing its text content as per the data in Table 2.

Category	**No. of PP URLs**
Non-English browser extension privacy policies	3,855
Browser extension privacy policies with no text feature	206
Browser extension privacy policies with Network error	2,090
Browser extension privacy policies for further processing	13,205
Total	19,356

Table 2: Processed category of Privacy Policies

Based on Table 2, we have 13,205 clean and active privacy policies, which can be used for further processing.

6 Challenges faced

During our course of study, we faced a few challenges. One of our challenges was getting the list of all workable URLs. We crawled the Google Chrome web store and got a list of browser extension URLs, but 5,412 URLs showed "Network Error" and could not be processed. On running the "Error set" again the second time, 36 URLs gave output and none after that. Nevertheless, these non-working URLs are still in the web store and can tamper with the result if not looked into it carefully.

Our next challenge was "network issue" with privacy policy URLs. The program was run three times with a timeout after five tries, i.e., 15 times in total. We are unsure if those URLs would have worked if we had increased the number of attempts before the timeout.

7 Conclusion

Data privacy and management of data are two essential concepts. Their importance is highlighted by the fact that most countries have distinct regulations concerning the protection and management of data. Like GDPR (General Data Protection Regulation) applies to the European Union's citizens, the US does not have a Federal law concerning the same. It follows the CCPA (California Consumer Privacy Act) for citizens of California, Virginia Consumer Data Protection Act (CDPA) for citizens of Virginia, etc., with laws based on particular groups of citizens like the Health Insurance Portability and Accountability Act (HIPAA) for health-related, Children's Online Privacy Protection Act (COPPA) to protect children's rights, etc. 157 countries have their own data privacy laws and guidelines as of March 2022 [27]. Most of the data protection and management policies are influenced substantially by the EU's GDPR.

According to the policies, the users have the right to know which of their data is being used, stored or given to third parties. The companies, or, in our case, the browser extensions, make these details available to users by publishing privacy policies on their websites.

This document concludes by emphasizing the significance of data privacy and management. It acknowledges the existence of regulations like GDPR and highlights the

influence of such policies on data protection practices globally. The study specifically focuses on whether browser extensions comply with essential data protection guidelines, primarily publishing privacy policies. The document successfully addresses RQ1 that not all browser extensions adhere to the practice of providing a separate privacy policy. The study's results contribute to quantifying the percentage of browser extensions having external privacy policies. It states that only 27% (as of May 2023) of browser extensions from the Google Chrome Web Store inform users about their data collection and management practices through a separate privacy policy.

Additionally, we found that multiple extensions share the same privacy policies, answering our RQ2 by revealing the frequency of shared privacy policies among browser extensions. As noticed during our analysis 3,773 privacy policies were used for 14,579 browser extensions in May 2023. This analysis emphasizes the need for closer scrutiny and examination of the content and specificity of these policies.

The results of our study contribute to the understanding of the current state of privacy practices among browser extensions in the Google Chrome Web Store and suggest areas for improvement in transparency and user awareness. The findings suggest opportunities for improvement in the communication of data practices by browser extensions. Developers may need to consider enhancing their transparency measures and tailoring privacy policies to reflect the specific data-handling practices of each extension. The study also contributes to the ongoing discourse on data privacy by shedding light on the state of affairs within the Google Chrome Web Store. It underscores the importance of continual scrutiny and improvement in privacy practices.

Overall, this report contributes to the increasing number of studies on privacy policies, specifically with browser extensions and showcases the potential of NLP techniques in analyzing text data.

References

[1] Ryan Amos, Gunes Acar, Eli Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. Privacy policies over time: Curation and analysis of a million-document dataset. In *Proceedings of the Web Conference 2021*, pages 2165–2176, 2021.

[2] Benjamin Andow, Samin Yaseer Mahmud, Wenyu Wang, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Tao Xie. {PolicyLint}: investigating internal privacy policy contradictions on google play. In *28th USENIX security symposium (USENIX security 19)*, pages 585–602, 2019.

[3] Benjamin Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Serge Egelman. Actions speak louder than words:{Entity-Sensitive} privacy policy and data flow analysis with {PoliCheck}. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 985–1002, 2020.

[4] Sruthi Bandhakavi, Nandit Tiku, Wyatt Pittman, Samuel T King, P Madhusudan, and Marianne Winslett. Vetting browser extensions for security vulnerabilities with vex. *Communications of the ACM*, 54(9):91–99, 2011.

[5] Adam Barth, Adrienne Porter Felt, Prateek Saxena, and Aaron Boodman. Protecting browsers from extension vulnerabilities. *Google Research Publications*, 2010.

[6] Jaspreet Bhatia, Travis D Breaux, Joel R Reidenberg, and Thomas B Norton. A theory of vagueness and privacy risk perception. In *2016 IEEE 24th International Requirements Engineering Conference (RE)*, pages 26–35. IEEE, 2016.

[7] Guido Boella, Luigi Di Caro, Michele Graziadei, Loredana Cupi, Carlo Emilio Salaroglio, Llio Humphreys, Hristo Konstantinov, Kornel Marko, Livio Robaldo, Claudio Ruffini, et al. Linking legal open data: breaking the accessibility and language barrier in european legislation and case law. In *Proceedings of the 15th International conference on artificial intelligence and law*, pages 171–175, 2015.

[8] Jasmine Bowers, Bradley Reaves, Imani N Sherman, Patrick Traynor, and Kevin Butler. Regulators, mount up! analysis of privacy policies for mobile money services. In *Thirteenth symposium on usable privacy and security (SOUPS 2017)*, pages 97–114, 2017.

[9] Jasmine Bowers, Imani N Sherman, Kevin RB Butler, and Patrick Traynor. Characterizing security and privacy practices in emerging digital credit applications.

In *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, pages 94–107, 2019.

[10] Brave. What are browser extensions, and are they safe?, 2023. URL https://brave.com/learn/what-are-web-browser-extensions/.

[11] Carolyn A Brodie, Clare-Marie Karat, and John Karat. An empirical study of natural language parsing of privacy policy rules using the sparcle policy workbench. In *Proceedings of the second symposium on Usable privacy and security*, pages 8–19, 2006.

[12] Duc Bui, Brian Tang, and Kang G. Shin. Detection of inconsistencies in privacy practices of browser extensions. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 2780–2798, 2023. doi: 10.1109/SP46215.2023.10179338.

[13] Nicholas Carlini, Adrienne Porter Felt, and David Wagner. An evaluation of the google chrome extension security architecture. In *21st USENIX Security Symposium (USENIX Security 12)*, pages 97–111, 2012.

[14] Quan Chen and Alexandros Kapravelos. Mystique: Uncovering information leakage from browser extensions. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1687–1700, 2018.

[15] Giuseppe Contissa, Koen Docter, Francesca Lagioia, Marco Lippi, Hans-W. Micklitz, Przemysław Pałka, Giovanni Sartor, and Paolo Torroni. Claudette meets gdpr: Automating the evaluation of privacy policies using artificial intelligence. *SSRN Papers*, 2018. doi: http://dx.doi.org/10.2139/ssrn.3208596. URL https://ssrn.com/abstract=3208596.

[16] Elisa Costante, Jerry den Hartog, and Milan Petković. What websites know about you: Privacy policy analysis using information extraction. In *International Workshop on Data Privacy Management*, pages 146–159. Springer, 2012.

[17] Lorrie Faith Cranor, Pedro Giovanni Leon, and Blase Ur. A large-scale evaluation of us financial institutions' standardized privacy notices. *ACM Transactions on the Web (TWEB)*, 10(3):1–33, 2016.

[18] Michael Curtotti and Eric McCreath. A right to access implies a right to know: An open online platform for research on the readability of law. *J. Open Access L.*, 1: 1, 2013.

[19] Noel Alvarez Domingo. Harnessing the power of chrome extensions and the benefits of chrome api, 2023. URL https://www.linkedin.com/pulse/harnessing-power-chrome-extensions-benefits-api-noel-%C3%A1lvarez/.

[20] Clover Dynamics. Browser extension: Everything you need to know, 2023. URL https://www.cloverdynamics.com/blogs/browser-extension-everything-you-need-to-know.

[21] Manuel Egele, Christopher Kruegel, Engin Kirda, and Giovanni Vigna. Pios: Detecting privacy leaks in ios applications. In *NDSS*, volume 2011, page 18th, 2011.

[22] Benjamin Eriksson, Pablo Picazo-Sanchez, and Andrei Sabelfeld. Hardening the security analysis of browser extensions. In *The 37th ACM/SIGAPP Symposium On Applied Computing SAC 22*, 2022. doi: 10.1145/3477314.3507098.

[23] Christine Fisher. 85 percent of chrome apps and extensions lack a privacy policy, 2019. URL https://www.engadget.com/2019-02-22-chrome-app-extension-security-flaws.html?guce_referrer=aHR0cHM6Ly9pYXBwLm9yZy8&guce_referrer_sig=AQAAACgMZTHBqA5DY0WVHV21_Ohv9h4ZMFCiG0LzefyvANZZkyFIdwDbaodPMoPt5E4cLWil-vvFkBrJEqPpl-9DezlHH3j4btnIfKDHmPdjHQ7Czjk8cwDw-5n8m_B-ByxAGkSAbVOPHz2mPxY6CK0J7QXFbHBoV3NPUe3u3NRaRmz_.

[24] Google. Google chrome web store developer agreement, 2021. URL https://developer.chrome.com/docs/webstore/terms/#privacy.

[25] Google. Set chrome app & extension policies, 2023. URL https://support.google.com/chrome/a/answer/7532015?hl=en.

[26] Peter Grad. Researchers issue warning over chrome extensions that access private data, 2023. URL https://techxplore.com/news/2023-09-issue-chrome-extensions-access-private.html.

[27] Graham Greenleaf. Now 157 countries: Twelve data privacy laws in 2021/22, 2022. URL https://ssrn.com/abstract=4137418.

[28] Arjun Guha, Matthew Fredrikson, Benjamin Livshits, and Nikhil Swamy. Verified security for browser extensions. In *2011 IEEE symposium on security and privacy*, pages 115–130. IEEE, 2011.

[29] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 531–548, Baltimore, MD, 2018. USENIX Association. ISBN 978-1-939133-04-5. URL https://www.usenix.org/conference/usenixsecurity18/presentation/harkous.

[30] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G Shin, and Karl Aberer. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 531–548, 2018.

[31] Chris Hoffman. Browser extensions are a privacy nightmare: Stop using so many of them, 2017. URL https://www.howtogeek.com/188346/why-browser-extensions-can-be-dangerous-and-how-to-protect-yourself/.

[32] ICO. Individual rights, 2023. URL https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/.

[33] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 471–478, 2004.

[34] Michael Kan. Most chrome extensions have no listed privacy policy, 2019. URL https://uk.pcmag.com/news/119783/most-chrome-extensions-have-no-listed-privacy-policy.

[35] Alexandros Kapravelos, Chris Grier, Neha Chachra, Christopher Kruegel, Giovanni Vigna, and Vern Paxson. Hulk: Eliciting malicious behavior in browser extensions. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 641–654, 2014.

[36] Ankit Kariryaa, Gian-Luca Savino, Carolin Stellmacher, and Johannes Schöning. Understanding users' knowledge about the privacy and security of browser extensions. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 99–118. USENIX Association, 2021. ISBN 978-1-939133-25-0. URL https://www.usenix.org/conference/soups2021/presentation/kariryaa.

[37] Masha Komnenic. 9 reasons why you need a privacy policy, 2021. URL https://termly.io/resources/articles/why-you-need-a-privacy-policy/.

[38] Juniper Lovato, Philip Mueller, Parisa Suchdev, and Peter S Dodds. More data types more problems: A temporal analysis of complexity, stability, and sensitivity in privacy policies, 2023. URL https://doi.org/10.1145/3593013.3594065.

[39] Sunil Manandhar, Kaushal Kafle, Benjamin Andow, Kapil Singh, and Adwait Nadkarni. Smart home privacy policies demystified: A study of availability, content, and coverage. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 3521–3538, 2022.

[40] Aleecia M McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *Isjlp*, 4:543, 2008.

[41] George R Milne, Mary J Culnan, and Henry Greene. A longitudinal assessment of online privacy notice readability. *Journal of Public Policy & Marketing*, 25(2): 238–249, 2006.

[42] Conor Murray. U.s. data privacy protection laws: A comprehensive guide, 2023. URL https://www.forbes.com/sites/conormurray/2023/04/21/us-data-privacy-protection-laws-a-comprehensive-guide/.

[43] University of Wisconsin-La Crosse. Web browser - what is a web/internet browser?, 2023. URL https://kb.uwlax.edu/page.php?id=89012.

[44] Ehimare Okoyomon, Nikita Samarin, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, Irwin Reyes, Álvaro Feal, Serge Egelman, et al. On the ridiculousness of notice and consent: Contradictions in app privacy policies. In *Workshop on Technology and Consumer Protection (ConPro 2019), in conjunction with the 39th IEEE Symposium on Security and Privacy*, 2019.

[45] Marta Otto. Regulation (eu) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation–gdpr). In *International and European Labour Law*, pages 958–981. Nomos Verlagsgesellschaft mbH & Co. KG, 2018.

[46] Harshvardhan J Pandit, Declan O'Sullivan, and Dave Lewis. Queryable provenance metadata for gdpr compliance. *Procedia Computer Science*, 137:262–268, 2018.

[47] Callum Pilton, Shamal Faily, and Jane Henriksen-Bulmer. Evaluating privacy - determining user privacy expectations on the web. *Computers '&' Security*, 105: 102241, 2021. ISSN 0167-4048. doi: https://doi.org/10.1016/j.cose.2021.102241. URL https://www.sciencedirect.com/science/article/pii/S0167404821000651.

[48] Privacy Policies. What's data privacy law in your country, 2023. URL https://www.privacypolicies.com/blog/privacy-law-by-country/.

[49] Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmeck, and Norman Sadeh. Identifying the provision of choices in privacy policy text. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2774–2779, 2017.

[50] Rocky Slavin, Xiaoyin Wang, Mitra Bokaei Hosseini, James Hester, Ram Krishnan, Jaspreet Bhatia, Travis D Breaux, and Jianwei Niu. Toward a framework for

detecting privacy policy violations in android application code. In *Proceedings of the 38th International Conference on Software Engineering*, pages 25–36, 2016.

[51] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N Cameron Russell, et al. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1330–1340, 2016.

[52] Lu Yang, Xingshu Chen, Yonggang Luo, Xiao Lan, and Li Chen. Purext: automated extraction of the purpose-aware rule from the natural language privacy policy in iot. *Security and Communication Networks*, 2021:1–11, 2021.

[53] Kinza Yasar. Are browser extensions really safe?, 2021. URL https://www.makeuseof.com/are-browser-extensions-really-safe/.

[54] Le Yu, Tao Zhang, Xiapu Luo, and Lei Xue. Autoppg: Towards automatic generation of privacy policy for android applications. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, pages 39–50, 2015.

[55] Le Yu, Xiapu Luo, Xule Liu, and Tao Zhang. Can we trust the privacy policies of android apps? In *2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 538–549. IEEE, 2016.

[56] Bin Zhao and Peng Liu. Private browsing mode not really that private: Dealing with privacy breach caused by browser extensions. In *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, pages 184–195, 2015. doi: 10.1109/DSN.2015.18.

[57] Sebastian Zimmeck and Steven M. Bellovin. Privee: An architecture for automatically analyzing web privacy policies. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 1–16, San Diego, CA, 2014. USENIX Association. ISBN 978-1-931971-15-7. URL https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/zimmeck.

[58] Sebastian Zimmeck and Steven M Bellovin. Privee: An architecture for automatically analyzing web privacy policies. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 1–16, 2014.

[59] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven Bellovin, and Joel Reidenberg. Automated analysis of privacy requirements for mobile apps. In *2016 AAAI Fall Symposium Series*, 2016.