

An Analysis Of Personal Data Shared To Third Parties By Web Services

Kirti Patil¹, Anubha Shahane², Shantanu Londhe³, Dr. Prashant S. Dhotre⁴

^{1,2,3,4}, Department of Computer Engineering, Sinhgad Institute of Technology and Science, Pune - 411041, India

Abstract - Today, internet has become a significantly dependent source for people, providing huge number of services for their day to day living. Varieties of web services are being used by a large number of people. However, there is a growing concern of the personal information privacy. The objective of this paper is to study and analyze the information acquired by web services from the users in Indian context. Study and analysis of more than 50 most used web services in India is made with reference to their privacy policy documents. The studied information is further visually represented using pie charts in order to give the user a clear idea of up to what extent his personal information is accessed by web services. Thus, the awareness regarding information privacy is brought to light through this survey.

1. INTRODUCTION

Have we ever thought how internet has gradually brought transformation into human lives? Earlier everything was used to be done without internet, and now almost everything is done through internet. The growth of internet is constantly expanding and it is quite true that we are enormously dependent on it. From booking a ticket to buying clothes, from finding a job to finding a spouse and from maintaining contact with friends to internet banking, internet has turned out to be the primary medium. But in this world where people are using internet in tremendous amount, how many of us have an idea of how our personal data is acquired, processed and exposed, without our consent? [1] While talking about use of internet on smart phones, in a survey of more than 2000 Americans, the research showed that 54% of users chose not to install an application after knowing how much personal information is actually acquired by the application. 30% of users uninstalled the application present on their phone because they realized that it collected their personal information that they were not willing to share [5]. In another survey, 70% of the users said that they would “definitely not allow” a service provider to use their location for their advertisement purpose. [6] Also a survey showed that 60% of users were very upset when they knew the app shares their location to advertisers. [7] Information privacy is something which most users are not aware of, but still is an issue to be concerned of. Whenever we start to use any web service, we are provided with privacy policy. How many of us read that policy document properly? And even if we read the document, how can we assure that everything mentioned is correct and followed properly by the web services. Also, many times we receive SMS or email from random senders related to loan, offers or any other advertisement. How many of us know from exactly where do they acquire our phone number or email id? The most common web services we use are the major providers of our personal data to those third parties. This paper will clarify most of these doubts.

2. TERMS AND MOTIVATION

2.1 Problem statement

“How to visualize the users’ personal information shared by service providers to third-parties? “

2.2 Definitions, Acronyms and abbreviations

Privacy refers to information shared with visiting sites, how that information is used, who that information is shared with, or if that information is used to track users [8].

Personally identifiable information(PII) is information that identifies you personally, such as your name, postal address, telephone number, or email address [9].

Non-Personal information can be technical information or it can be demographic information, such as your age, gender, ZIP code or other geo location data, or interests. Non-personal information does NOT identify you personally.

Third party cookies - When users visit APA websites third-party cookies (such as Google advertising cookies) may be sent to their computers by APA's advertisers or service providers.

They may use the information obtained from these cookies:

- To track users browsing across multiple websites.
- To build profiles of users' Web surfing.
- To target advertisements that may be of particular interest to users.

Behavioral information is the capture of interaction, click, engagement, movement through a website, actions taken via email marketing across the buyer journey, user movement through digital content and more.

An existing system named as "Disconnect.me" is an extension provides information about third parties when user visits a particular website. This system has provided a way to retrieve the information about third parties. This is helpful for increasing an awareness about privacy among users.

3. SURVEY AND PROPOSED METHODOLOGY

The aim of the survey mentioned in this paper is to analyze and understand different type of users' data collected by different web services. Also, this paper focuses on awareness regarding user's privacy of information. This survey was done by the authors.

3.1 Method of survey

The following method is followed while carrying out the survey. Initially, 52 most commonly used web services in Indian context were taken into consideration. This was done based on the Alexa rankings of the web services. This ranking is as per the visits made by Indians. Further, the privacy policy documents of each service were studied thoroughly by the authors. Study was made regarding what type of users' personal data is collected by those services. For example, service 'xyz' has mentioned in their privacy policy that they are collecting following data First name, email id, phone number, IP address. They are not collecting - location, password. And it has not mentioned if they collect - Birth date, banking details. Information of third parties are collected through existing system and using 'href' of privacy policy. Data collected by third parties are also studied with help of web service's privacy policy and third party's privacy policy.

3.2 Data Analysis

The studied data was collected in a spreadsheet. In all 83 attributes like first name, last name, email id etc. were formulated. With reference to what is mentioned in the privacy policy, in front of each web service, it was specified that if they collect a particular data attribute or not or is not mentioned. ('Yes', 'No', 'Not specified') This information was then visualized using a tool 'Tableau' which is a business intelligence tool [www.tableausoftware.com]. This tool enables us to visualize data. This is a very simple-to-use tool which also supports drag and drop facility. Pie charts were formed depicting the data. Further, the data attributes were classified into - PII (Personally identifiable information), Behavioral, Device information and other information. Finally, this processed information was represented to user along with suitable description. A website providing selection of category of web service and then selection of third party is built. User friendly visualization is made. Information about third parties and data shared to them is displayed.

3.3 Architecture and design

The design methods chosen for the development of the systems had to be high on reliability when classification constraints are considered. The extraction method had to be well versed which could be done well on a client-server system. For the project to work on real time on most of the functioning and allowing websites we had to trade a little bit on accuracy to make the scale larger and widely acceptable to the users worldwide.

This design provides an idea about system to be developed. When user visits a web site he/she will select a category of web service (e.g. Educational, shopping etc.). After selection of category selection of web service is done. Web service may consist of various categories of third parties which are associated with it. User will select a category of third party

(e.g. Analytics, social etc.). From the given list of third parties one will be selected. And then information of selected third party is displayed along with the data getting shared to them by web services. For this purpose, all data are collected from privacy policies of web services. All policies are read and data collection which is mentioned in their privacy policy is collected. Third party detection is done using existing system named as “Disconnect.me” and by reading privacy policies using “href”. All collected data are stored in excel sheet and visualized with help of JSON, JavaScript.

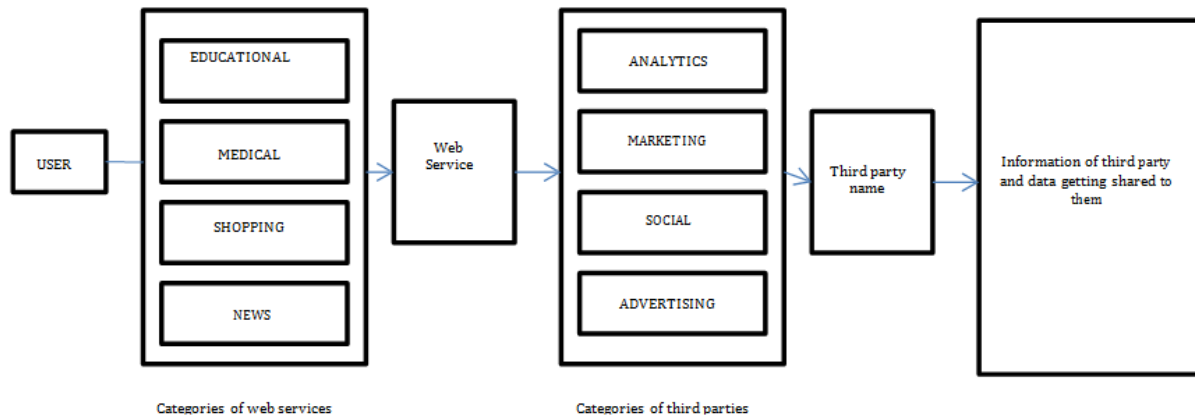


Fig-1: System Architecture

4. RESULTS AND DISCUSSION

In the following, we have added a visualization of website. Home page showing the categories of web services (e.g Shopping/Business, news). User has to choose a category of required webservice first. There is a search button for searching particular web service.

After selecting particular category all names of web services get displayed. User has to select one of required and then category of third party need to be selected. After selection of category, name of third parties are displayed. Then required information and shared data are visualized.

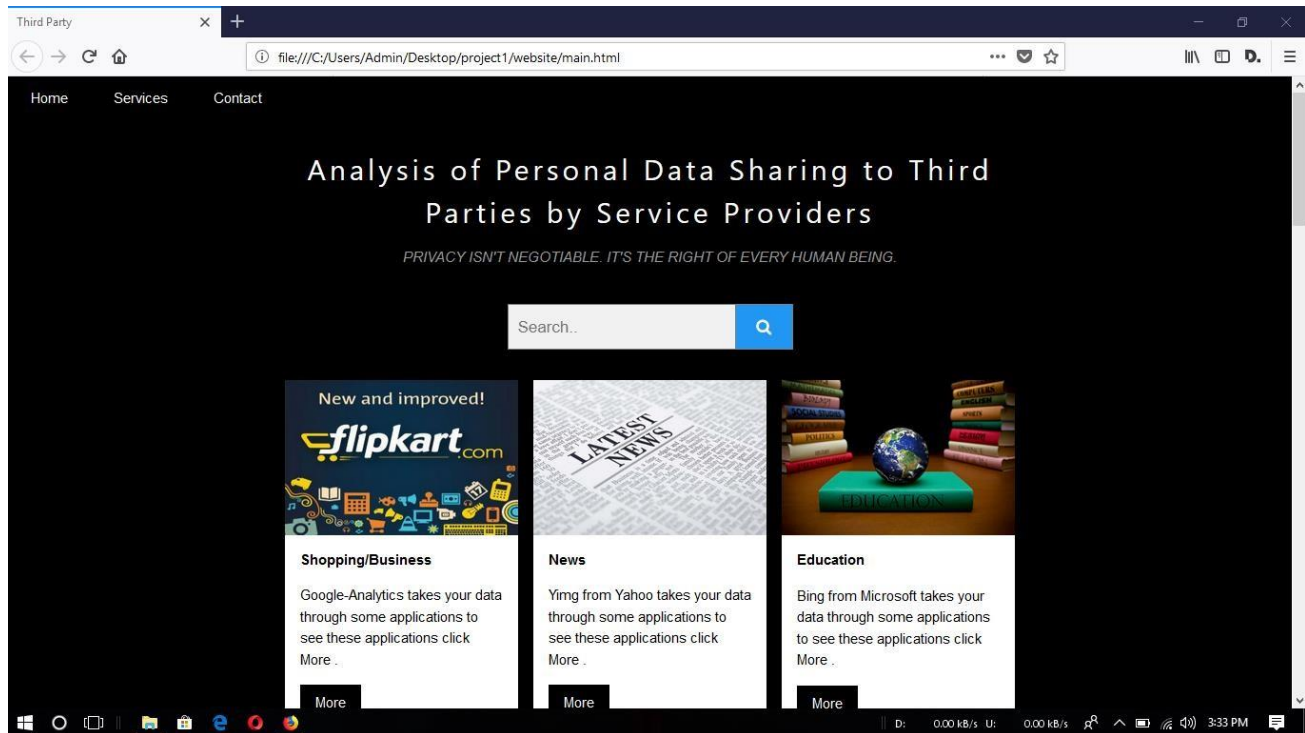


Fig-2: Home page visualization

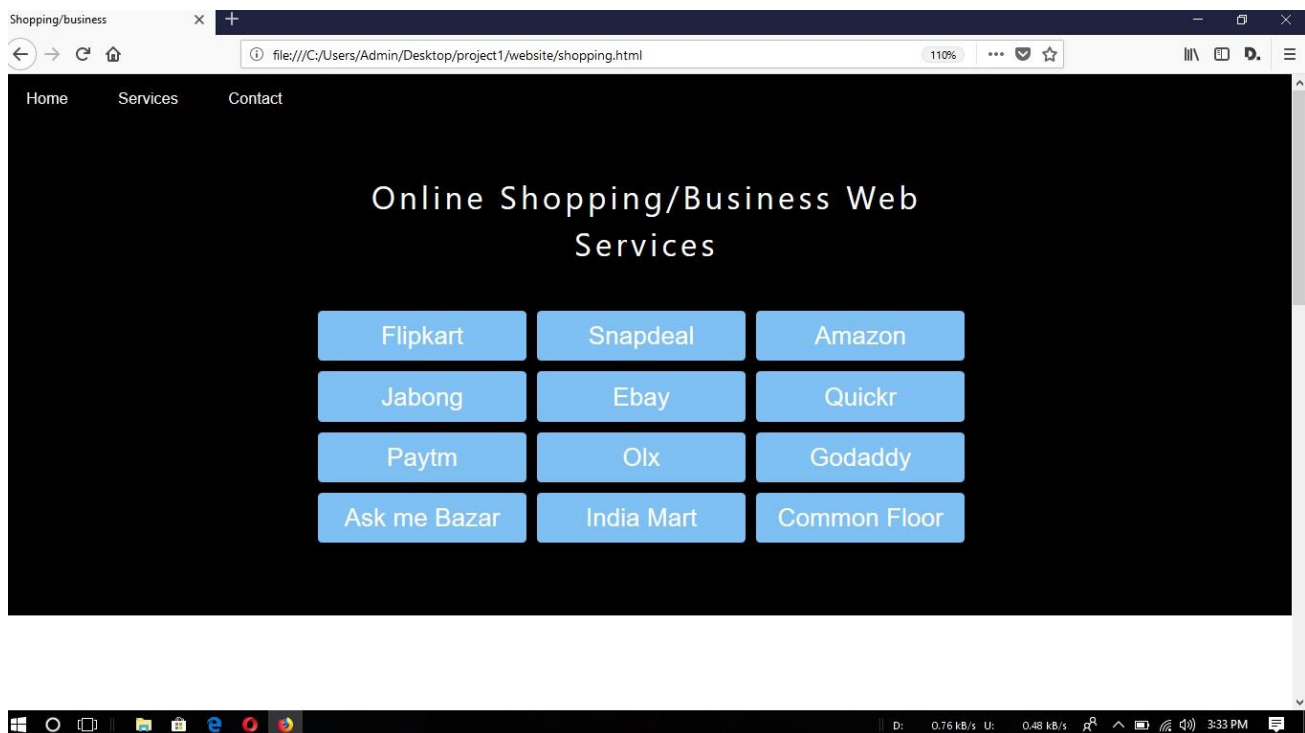


Fig-3: Web service names

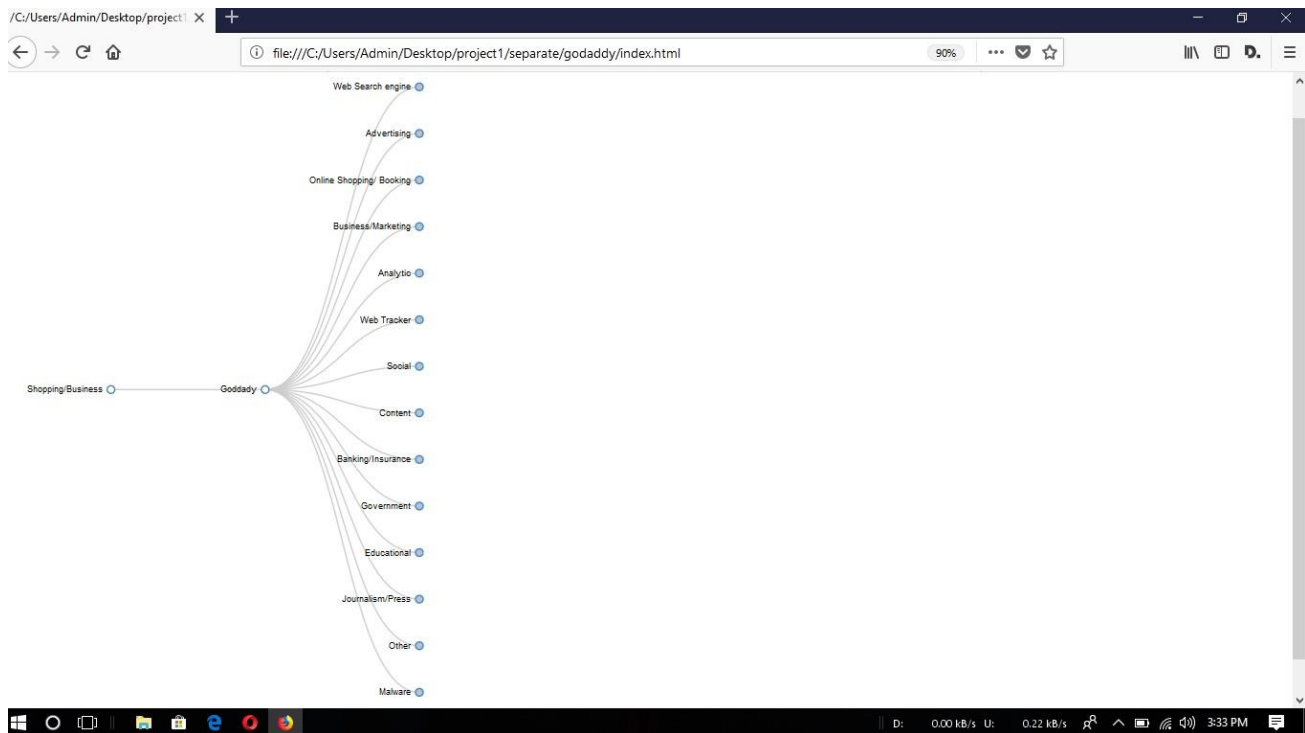


Fig-4: Tree visualization

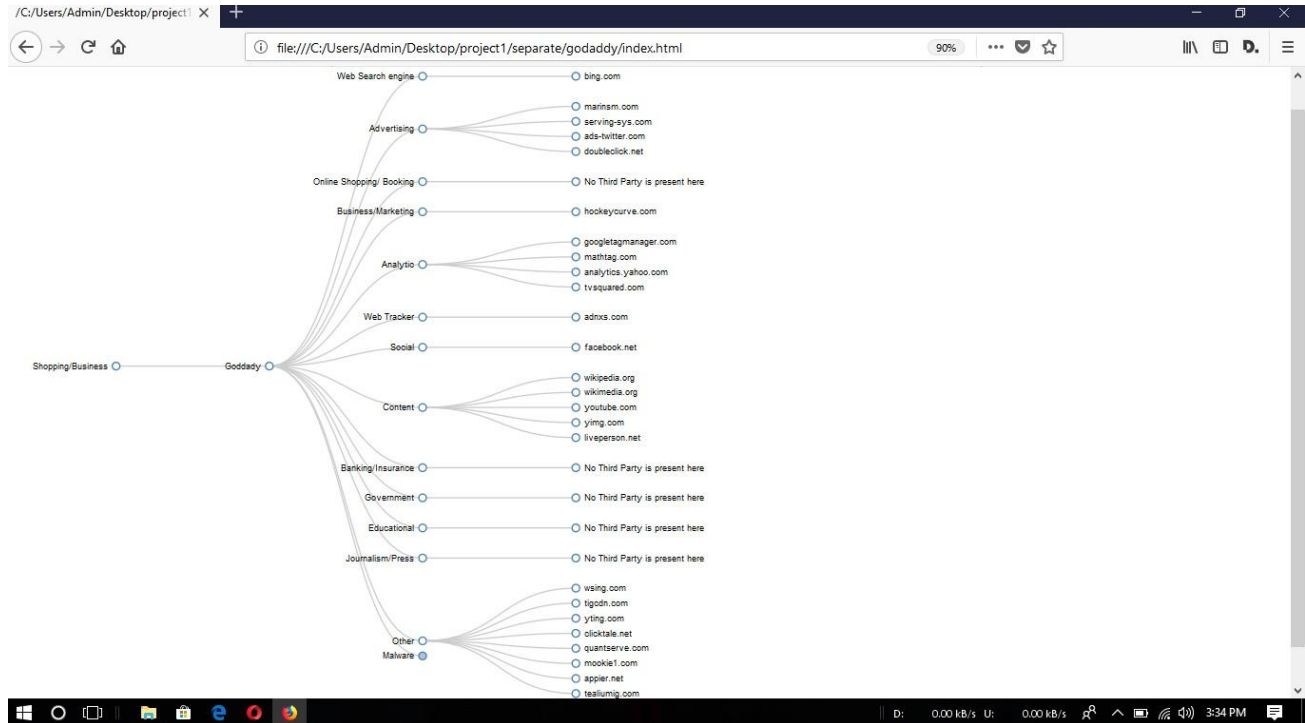


Fig-5: Data shared to third parties

5. CONCLUSION

In this project, we have provided an efficient way to understand the privacy and data sharing to third parties by web services. Considering privacy and personal data, web services are classified into categories. User friendly visualization helps to search the web services and to select a particular third party along with its category about which information is to be obtained.

Analysis of privacy policies of web services and privacy policies of third parties are done. With the help of existing technology some of the third parties are captured. Further studies on the third parties are done for understanding their privacy policies and acquiring the information about them. After this analysis data sharing to third parties are studied along with web service's data mentioned in privacy policies.

Considering user's perspective, a potent visualization of analyzed and studied data is done. This will help users to understand the privacy policies and data sharing to third parties.

Provision of information of third parties and data

getting shared to them through web services gives an idea about privacy. For some of web services and their associated third party's data sharing or information are unavailable, those will be done further with more analysis of particular web services.

6. REFERENCES

- [1] A Survey of Privacy Awareness and Current Online Practices of Indian Users, Dhotre, Prashant Shantaram; Olesen, Henning.
- [2] Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps Jinyan Zang, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney.
- [3] Privacy Scoring and Users' Awareness for Web Tracking Asma Hamed, Hella affel-Ben Ayed CRISTAL Lab. ENSI, University of Manouba Tunis, Tunisia Hella Kaffel-Ben Ayed Faculty of Science of Tunis University of Tunis El Manar Tunis, Tunisia.
- [4] Making your privacy public, Kamala D. Harris, Attorney General California Department of Justice.
- [5] Boyles J, Smith A, Madden M. Privacy and Data Management on Mobile Devices. Pew Research Center. September 5, 2012. http://www.pewinternet.org/files/oldmedia/Files/Reports/2012/PIP_MobilePrivacyManagement.pdf
- [6] Urban J, Hoofnagle C, Li S. Mobiles Phones and Privacy. BCLT Research Paper Series. July 10, 2012. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103405
- [7] Felt A, Egelman S, Wagner D. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices. October 19, 2012. <http://dl.acm.org/citation.cfm?id=2381943>
- [8] <https://www.computerhope.com> Dictionary P-Definitions
- [9] <https://virtualinc.com/legal-privacy-identifiable-information>
- [10] <https://projectinnovation.org/page.php?name=privacy-policy>
- [11] <https://searchcio.techtarget.com/definition/data-privacy-information-privacy>
- [12] <https://www.techopedia.com/definition/10380/information-privacy>
- [13] https://en.wikipedia.org/wiki/Information_privacy
- [14] <https://searchcio.techtarget.com/definition/data-privacy-information-privacy>