



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 5)

Available online at: www.ijariit.com

Collection of data through cookies and smart devices– A case study

Garima Kaushik

garima95.gk@gmail.com

Rishabh Prakash

rishabhprakash01@gmail.com

ABSTRACT

Data is everywhere. Ubiquitously, companies and governments are collecting data. One simple visit to the World Wide Web leaves a wide digital footprint. One Google search about the best mutual funds in the market and you will be bombarded with mutual fund advisories anywhere you visit. The lines between the online world and the offline world are blurring, as the number of metadata increases. This paper looks at certain ways as to how our civilization is becoming smarter, with every device around us right from the mobile phone to electric bulb, becoming smarter with each passing day. This paper looks at the origins of the collection of metadata, as well as their use in customizing the world as per our needs and wants.

Keywords— Big Data, Data Analytics, Business Analytics, Cookies, Privacy, Statistics

1. INTRODUCTION

The current world population exceeds 7.6 billion [1] and over 4 billion people are connected to the internet. [2] As a result of this technological revolution, millions of people are generating a massive amount of data every second. This data is too big, too hard and too fast for the conventional database system to process the data. And this data is known as Big Data. [3] The features of Big Data of big data can be characterized by 5V, namely, Volume, Velocity, Variety, Veracity, and Value. [4]



Fig. 1: Big Data

1.1 Why data is so important?

Information is the currency of the 21st century. [5] In this world of interconnected systems, the more companies know about their customer, the better will they be positioned to ensure better sales and prevent abuse of their systems and platforms.

Creating a customized experience for the user is very important because a company is only as valuable as its nuisance value. Nuisance value is a term coined to mean the hardship the customer will have to face if he stops using the company's products altogether. And nuisance value has a direct correlation to market share and market dominance. [6]

The clearest picture is that of Google. It is very difficult to move away from the products and services offered by Google. As of Nov 12, 2015, according to built with, a technology platform which checks what is the underlying machinery, behind this highly operative websites, almost 69.5% of Quantcast top 1000 websites use Google Analytics. [7] Just imagine how difficult is it for an ordinary person to go about steering clear to 70 out of every 100 websites on the internet, at the same time forgoing excellent email, maps, and search service? Terribly difficult, and probably borderline impossible. Hence, Google has a dominant position in the search market, almost a monopoly in the maps market (after the spectacular failure of the maps application of its biggest competitor), and has built an all-encompassing ecosystem.

As of September 2018, close to 70% of internet browser market share is controlled by Google Chrome. [8] And that is not all. Even the other browsers like Vivaldi, Opera, and Chromium are based on the same engine “Blink” [9]. Hence, even moving from Chrome to Opera is not going to alter anything in terms of moving away from a Google Ecosystem.

Table 1: Data Collection Graph

Source	Quantum of Data
YouTube [10]	YouTube has over a billion users almost one-third of all people on the Internet each day those users watch a billion hours of video, generating billions of views. This statistic shows a timeline with the worldwide number of monthly active Facebook users from 2008 to 2018. As of the second quarter of 2018, Facebook had 2.23 billion monthly active users. In the third quarter of 2012, the number of active Facebook users had surpassed one billion, making it the first social network ever to do so. Active users are those which have logged in to Facebook during the last 30 days. Furthermore, as of the previous quarter the social network had 1.74 billion mobile MAU. The platform is also the most popular social network worldwide.
Twitter [11]	Total Number of Monthly Active Twitter Users: 335 million (source) Last updated: 9/18/18. Total Number of Tweets sent per Day: 500 million. Last updated: 6/24/18. Percentage of Twitter users on Mobile: 80% Last updated: 6/24/18. Number of Twitter Daily Active Users: 100 million
Facebook [12]	This statistic shows a timeline with the worldwide number of monthly active Facebook users from 2008 to 2018. As of the second quarter of 2018, Facebook had 2.23 billion monthly active users. In the third quarter of 2012, the number of active Facebook users had surpassed one billion, making it the first social network ever to do so. Active users are those which have logged in to Facebook during the last 30 days. Furthermore, as of the previous quarter the social network had 1.74 billion mobile MAU. The platform is also the most popular social network worldwide.
Instagram [13]	As of June 2018, there are nearly 1 billion monthly active users. As of June 2018, there are nearly 500 million daily active users. The Like button is hit an average of 4.2 billion times per day
Quora [14]	Number of Quora visitors: more than 200 million monthly unique visitors 400,000 topics on Quora

2. AN INTRODUCTION TO COOKIES AND THEIR TYPES

2.1 Definition

A cookie is a small text file which is stored by a website you visit, on your local computer. Several times there are tick boxes saying “Remember me on this device” right on the login screen. What does it do? Well, clicking the box simply places a cookie on your computer stating that till the cookie expires, the computer must automatically log you into that particular service. [15]

2.2 Security and cookies

A cookie set by a web server can only be read by that particular server. Hence, the cookie poses no fundamental security risks. The webmasters have a lot of tools at their disposal to check the user activity across their websites, and cookies are just a tool to make it easier. Cookies in no way give access to user data on the hard disk, and by and large, they cannot be used to execute any malicious code on the host machine. It is just an identifier for that machine set by the server. [15]

2.3 Types of cookies

Two broad categories of cookies are session cookies and persistent cookies. Session cookies can be stored in a computer only for the duration of the user session on the website. Whereas the persistent cookies are stored for a long time, much after the session has ended.

2.3.1 Session cookies: They help the website remember the user, like for example the items added to shopping cart, or the activity done by the user, in the order of the pages opened by him, and also by providing the IP address other public information to the company. [16]

2.3.2 Persistent cookies: These are stored in the computer long after the session has ended. These give the website information about user’s login credentials, also they are used to uniquely identify the user across multiple sessions, saving long-term user preferences, and mainly drawing up a blueprint of user activities across the website. Since persistent cookies contain private information, they are usually encrypted by the web server, before storing them on the users’ computers. These cookies stay in the system for as long as they don’t expire or are not deleted. [16]

2.3.3 Further, cookies can also be classified into categories depending on the party which placed them on your system.

(i) **First party cookies:** These cookies are placed by the website itself, and can be easily identified by the domain name of the website. [17]

- (ii) **Third party cookies:** They are placed by advertisers and can be tagged to any advertising element on the website. These advertisers have an interest in determining which advertisement was shown to a particular user, how many times, and also the interaction of the user with their advertisement. [17]
- (iii) **Super cookies:** With increased talk about cookies and the data they store, there was a lot of pressure on web browsers to make it simple for users to delete their cookies. In most browsers ctrl + shift + delete should do the task. But there is a variety of cookie called the super cookie, which as the name suggests is a much more powerful version of the cookie. It is difficult to detect and delete, and in some cases, it cannot be deleted using basic browser cache cleaning methods, or even simple systems like a disk cleaning software. [18]

2.4 Cookie profiling

The major problem with cookies is a system called cookie profiling. Large advertisers get cookie information from several high volume websites. These cookies can then be placed together and the advertiser can collect a significant amount of data regarding the user's behavior, his choices, demographic and sometimes even his finances. At the very least this data can be put to use to provide targeted cross-platform ads to the customer. At times, these ads are not just cross-platform, they are also cross-system and cross services. [19]

For example, there is a good chance that advertisers glean through the cookies and through profiling they come up with information saying you are interested in say, cars. Then the same data will be used to show ads regarding the automobiles of your choice, even though the website on which the ad is placed has nothing to do with an automobile. You will also be targeted through calls and messages in the offline world, offering discounts and promotions on automobiles, and related services. It will almost feel as if the world is conspiring to ensure that you purchase an automobile or a related service, and guess what? In a way it actually is.

2.5 What information does this data hold about you?

The data analyzed through online and offline mediums can be broadly invasive and tell a lot about you and your habits and your lifestyle. In certain cases, the data might know what you close family members, and in rare cases, even you do not know.

For example, there was a Forbes article in 2012, mentioning Target, a major shopping chain in the United States of America, knew about the pregnancy of a teen girl, even before her immediate family. [20]

Data gathered through smart home lighting systems can tell a person about your sleeping / waking patterns. If carefully looked at, it can even predict the times at which a user is most likely to be or not to be at home, which can be a huge security concern. This will be very similar to the case where people posted on a social network before going on a vacation, which was read by miscreants, and led to a robbery of their home. These cases have become so severe, that insurance companies have started rejecting burglary claims in case the users posted about their vacation beforehand on the social network. [21]

3. CONCLUSION

The amount of data being generated every day is enormous. [22] Further, this has huge privacy implications. Hence, users must be very careful about the data they give out. And should take utmost care and good measures to ensure that there is no unwanted data leak. A good first step would be to begin reading the privacy policies and the data retention laws of your favorite online services. Be aware is to beware!

4. REFERENCES

- [1] Worldometers, "Real Time World Population", 2018, "<http://www.worldometers.info/world-population/>".
- [2] Internet World Statistics, Dec 2017, "<https://www.internetworldstats.com/emarketing.htm>"
- [3] Madden, S.: "From Databases to big data", IEEE Internet Computing 16(3), 4–6 (2012)
- [4] X. Jin et al., Significance and Challenges of Big Data Research, Big Data Research (2015), "<https://doi.org/10.1016/j.bdr.2015.01.00>"
- [5] Personal Information is the Currency of the 21st Century, "<http://allthingsd.com/20130507/personal-information-is-the-currency-of-the-21st-century/>"
- [6] Strategies for High Market-Share Companies, "<https://hbr.org/1975/11/strategies-for-high-market-share-companies>"
- [7] Quantcast Measurement Usage Statistics, Built With Trends, "<https://trends.builtwith.com/analytics/Quantcast-Measurement>"
- [8] Browser Market Share Worldwide, Stat Counter Global Stats, 2018, "<http://gs.statcounter.com/browser-market-share>"
- [9] Vivaldi, Opera, Chromium are based on the same engine "Blink", "[https://en.wikipedia.org/wiki/Vivaldi_\(web_browser\)](https://en.wikipedia.org/wiki/Vivaldi_(web_browser))"
- [10] YouTube, "YouTube statistics," 2018, "<https://www.youtube.com/yt/about/press/>"
- [11] Twitter, "Twitter statistics", 2018, "<https://www.omnicoreagency.com/twitter-statistics/>"
- [12] Facebook, "Facebook statistics", 2018, "<https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>"
- [13] Instagram, "Instagram statistics", 2018, "<https://adespresso.com/blog/instagram-statistics/>"
- [14] Quora, "Quora statistics", 2018, "<https://expandedramblings.com/index.php/quora-statistics/>"
- [15] Jonathan R. Mayer, John C. Mitchell: "Third-Party Web Tracking: Policy and Technology", Published in 2012 IEEE Symposium on Security and Privacy, Online ISSN: 1081-6011
- [16] Types of Cookies, Verisign, "https://www.verisign.com/en_GB/domain-names/online/implement/what-are-cookies/index.xhtml"
- [17] Types of Cookies, "<https://www.grahamcluley.com/whats-difference-first-third-party-cookies/>"
- [18] Super Cookies, "<https://searchsecurity.techtarget.com/definition/supercookie>"

- [19] Stefano Traverso, Martino Trevisan, Leonardo Giannantoni, Marco Mellia, and Hassan Metwalley: “Benchmark and comparison of tracker-blockers: Should you trust them?” Published in: 2017 Network Traffic Measurement and Analysis Conference (TMA), Electronic ISBN: 978-3-901882-95-1.
- [20] Forbes article, 2012, “<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>”.
- [21] Insurance claims rejected, <https://www.dailymail.co.uk/news/article-3051671/Holidaymakers-post-information-trips-Facebook-face-having-insurance-claims-rejected-home-targeted-burglars-away.html>.
- [22] Umar Ahsan, Abdul Bais: “A Review on Big Data Analysis and Internet of Things”, Published in 2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Electronic ISSN: 2155-6814.