International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V69I7P106

Volume 69 Issue 7, 43-47, July 2021

Original Article

At the Cost of Consumer's Privacy, Data Becomes the Most Expensive Digital Currency

Munjal Shah

Kellog School of Management, Northwestern University, Chicago, The United States of America.

Received Date: 03 June 2021
Revised Date: 07 July 2021
Accepted Date: 17 July 2021

Abstract *- In the IT industry, data growth is increasing with every passing day. The need of the day is to have the individual rights reserved to share or use the data as per the desired cost. The implications of the data submission were not fully realized until technological maturity was reached. The right to use and manipulate the data resides with the data owners and handlers: data aggregators, websites, socializing platforms, utility companies, brokers, etc. Had the data ownership been restricted to the individual whom the data identifies, the data giants would have to pay for the use of their data for any e-commerce, sales, or digital marketing purposes. Not forgetting to mention, the data payment should be something of monetary value and not some free service or offered discounts. This would break the monopoly of the data giants, and a poor individual may benefit by merely getting registered on a platform that asks for personal data access. This is a much-needed development required to revolutionize the data ownership industry. The data generators need to realize the business worth of their data asset. Every individual must safeguard his data, must ensure to reserve the legal right to have full authority to share his data and encash it at his own terms and conditions rather than offering it for free to already wealthy organizations, thus adding to their net worth.*

Keywords *- Data Currency, Machine Learning, Data Science, Data as a service, Consumer Privacy, Data Aggregators, Data Protectors.*

I. INTRODUCTION

Data collection and maintenance have been in practice for ages. With the data growth, the data storage means were modified over time. Let it be data collection for population census or for general vaccination records; logs and files were manually maintained by the government to keep the records. Over time, the data migrated to public domain archives. Initially, this data was handled and maintained by utility companies appropriately. Later, these companies compromised the people's privacy and sold the data access to online sales and marketing organizations.

Over the past few years, the data grew exponentially, coining a new term, "*data currency*", setting up a whole new business area in the market of digital currency. Google, Facebook, Twitter, Amazon, Microsoft are some of the giants amongst the list of data brokers and aggregators. While signing up for their so-called *free* accounts, they manipulate the consumer to accept their terms and conditions in which makes the consumer give up on his right to identity control. They claim to own the right to use the consumer's data.

The reports from 2011 told that Mark Zuckerberg earned $24 billion dollars that year through Facebook [5]. According to a survey, $23.4 million were spent on Mark's private air travel and security by Facebook [9]. Since Facebook accounts are free, how did Mark become a multi-billionaire?

While signing up, the consumer may be offered some incentive or discount offer as a cost for the use of data. The best example in this scenario is the free email service providers. They take the consumer's sign-up data and provide him with a *Free* account in return. They monetize the consumer's data by selling it to the advertising companies targeting the consumers of the email service. In addition to this, the data and demographics are also shared with digital marketers and sales at a hefty cost. The consumer is kept in shadow regarding the cost of the data compromise and running an email service. In the end, all that a consumer gets is a mere email account. This does not imply that every minute detail of the consumer is monitored and controlled by these reputable email providers, but the chances of individual vulnerability fairly increase.

In the article, the main issue under discussion is the lack of presence and enforcement of laws and by-laws safeguarding consumer's data from data aggregators. The concept of data privacy has been under debate by lawyers for many years now. Potential solutions for the issue are being proposed; consumers are constantly intimated about the potential data risk, data theft, and what preventive measures they can take to safeguard their data. However, there is no one to question the data aggregators and brokers who sell the data to multiple hands for multiple purposes without any proper verification. Neither are there any constraints for data accuracy check. The data can be manipulated, can be used against the individual, or even against the state in worse cases.

The article puts a shadow on the value of the consumers' data, the legal status of the data, future scope, and potential of the data and with whom the ownership of the data must reside. It further discusses some common issues and proposed solutions to the issues as well. There reside numerous personal and societal benefits with data ownership only if data potential is fully understood and the consumers strive and protest to revolutionize the data ownership.

II. DATA – THE NEW DIGITAL CURRENCY

"Data is the new oil" is a mantra across the globe. Like Oil in the 18th century, data is the most marketable asset of current times. With the growth of data, the volume of digital data multiplied was expected to increase by more than 40 times over the years 2013 to 2020 [8]. Those who have the data and know-how to extract and utilize it are getting huge rewards for this.

Today is the era of the digital economy, with data having the highest rate in the market as a currency. In the world of digital sales and marketing, the application of AI on the collected data helps to understand the interests and searching trends of the consumers. This helps the companies to target the consumers of their interest only. Data has become a key to smooth functionality for all local and government companies. But everything comes with a cost. Data phishing and theft do exist, which may result in lifetime assets loss of a consumer. For this reason, data is treated like gold in the digital currency with the highest monetary value.

III. DATA USAGE

Data is being collected from very early times and has a variety of purposes. However, the most popular use of the data is marketing. Initially, there were loyalty cards instigation to track user's purchases for the purposes of targeted marketing. However, there are other fair purposes of data collection as well. One such example is the census at the government level. This may help the government to make better economic policies. In changing populations, larger datasets better help in identifying the mass trends. Given below is a general communication model.

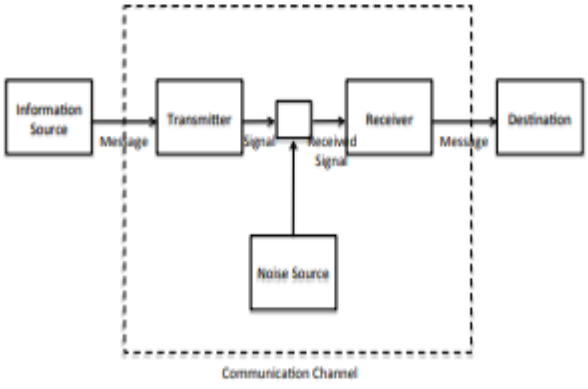

Fig. 1 Schematic Diagram of a General Communication System

A. Shannon's Model of Communication

In 1948, a mathematical model for data transmission was published by Shannon [6]. The mathematical modeling of the data is well beyond the scope of this article. So, we will be observing the schematic diagram of a Shannon Communication Model that he reproduced from a general communication model. In the simplest scenario, let us consider an individual being the data source and his messages being the data (e.g., address, age). When data is to be collected from multiple sources and has to be combined to generate an individual with a new identity, the situation becomes complex. To cope with the situation, Shannon's model was expanded.

B. Proposed Model

The model proposed that there are multiple sources of information, all feeding the data to the transmitter. These sources can include the user himself giving the data or the data aggregators or companies selling their data who have collected it before. The data from all such sources are combined to infer the complete profile of the user. The inferred data is most likely to be less accurate than the collected data and is considered noise. The noise is then combined with the collected data and then forwarded to the receiver. The receiver stores the data at its final destination. This data can be used as noise for later data collection processes for the same receiver as the data may contain noise already due to the procedure through which data was collected previously. This generates a feedback loop that continuously degrades the quality of the data asset [7].

The pictorial representation of Shannon's Communication Model is given below:

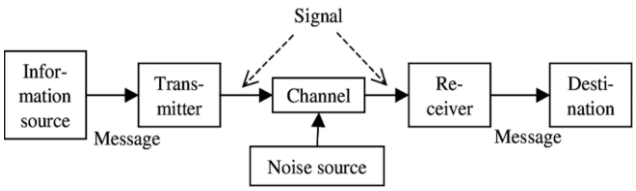

Fig. 2 Shannon's communication model

IV. PROBLEMS WITH DATA

With the collection of data, there are several associated issues. The main cause is due to the fact that data changes frequently, and errors may occur while data recording. Some issues are described below:

A. Data Accuracy

The data collected about individuals is often incorrect. The users are usually unaware of the fact that their data is being collected, let alone knowing who is handling the data and how to access the data for correction. Some brokers and data aggregators are offering mechanisms to access personal data for modification purposes. However, the majority of the users don't know about the data aggregators either. So, the accuracy of the data is always a question mark.

B. Data Timeliness

The data collected at a time may be accurate for that particular time but may be subject to change later on according to the circumstances. For example, when data was collected, the user had one phone number, but the user changed his number later on. So, the data was correct at the time of collection but changed later on.

C. Data Inference

The data about an individual may not be collected directly from the individual but from different sources. For example, a user's income may be inferred from his address if the house is rented or owned. Due to this, there is a varying level of accuracy achieved by data inference and results in a wide range of inaccurate data being populated.

V. WHO OWNS THE DATA

The concept of data currency starts dangling at the question, "Who is the owner of an individual's data?" Apparently, it seems like the data is owned by the individual himself. This is the biggest scam one can be fooled by. Some states have "laws of publicity" which prevent the public usage of an individual's data and identity without permission [4]. All the laws have a converging point which states "for commercial use of an individual's data, his permission is mandatory." Despite the fact that such laws exist, law enforcement seems missing. When registering on a retail website or on an email service provider, the individual has to give permission by signing an agreement with a single click, without which they do not let the individual register. There are fair chances that only diligent individuals go through those agreements. Millions of people sleepwalk into giving their identity to both well-reputed and dubious websites. The data is gathered from multiple sources and combined to come up with an entirely new profile of the individual. The authority of the individual's data resides nowhere with the individual.

It is a common misconception that a person has the authority over his personal data. The reality is quite in contrast with the authority of the data lying in the hands of the organizations which have the data in their servers and databases. In exchange for online facilities, these organizations demand data. The practice has become almost mandatory, and to register at any website, a user must enter his identity. This practice is justified by providing easy access to the website when approached next time. The authority behind the websites uses this data to expand their business at the cost of users' privacy by sales, direct marketing, or selling data access to third parties. The users are left with no choice than to permit the authorities to use their data else they cannot register.

Though data is also required in conventional businesses, for them, the data holds secondary value in comparison to the services they are providing. The businesses which see the data as their major asset face many difficulties in establishing ownership of the data.

VI. DATA THEFT

Data theft is the most common issue being faced worldwide. The data is not only sold by the data owners but is also stolen illegally by hackers. The hacked data is then sold in the black market for over billions of dollars.

Given below is an identity theft stats graph that shows identity theft over the year 2016 to 2020. The top five states based on the reported number of identity theft cases are Kansas, Rhode Island, Illinois, Nevada, and Washington DC [10].

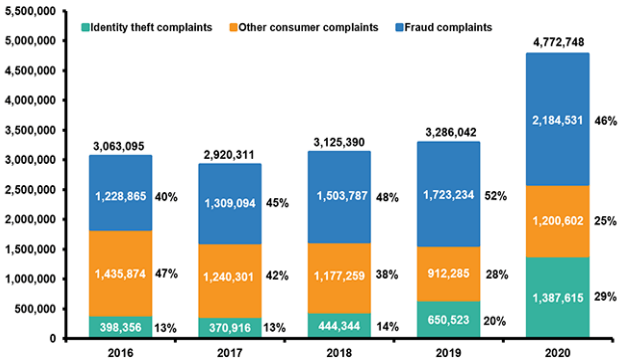

Fig. 3 Identity theft and fraud report 2016-2020

On 22nd January 2020, a database for customer support that contained over 280 million Microsoft customer records was left open on the internet [12].

On 14th April 2020, the data of over 500k Zoom accounts (teleconferencing) was found on the dark web for sale [13].

On 20th July 2020, sensitive data of 60,000 customers was exposed due to unsecured servers from family history search software company, Ancestory.com [14].

On 18th February 2021, the Automatic Fund Transfer Service of the California Department of Motor Vehicles (DMV) was hit by a massive ransomware attack. Their drivers got alert as they suffered a massive data breach [15].

VII. VALUE OF DATA

There is no well-defined price for the data. However, a close estimation can be done based on where the identity is sold, e.g., the black market. In July 2013, two Russian hackers were charged by the FBI for hacking more than 800,000 bank accounts resulting in the loss of millions of dollars [1]. The same year, a website named "Secure Works" released a study report on data values [3]. Some of the highlights from the report are given below:

- A fullz (complete identity: name, phone, email, address, ID number, passwords, birthdate, bank accounts with credentials) costs $25 in the United States and $40 in Europe.
- For the United States Visa/Mastercard, the price was

$4; for the UK or European credit cards, the price reached $8.

- $25-$100 for doxing. Doxing refers to stealing data from the target victim by infecting the victim's system with data-stealing trojan/via social engineering by hiring a hacker.
- The price for a bank account with $70,000 - $150,000 ranged up to $300.

In addition to this, the value of data can also be inferred from the cost (in terms of time and money both) victims have to pay for identity theft. Javelin Strategy and Research put forth that for a victim of identity theft, the average resolution time (ART) is 12 hours [2]. The equation for the final value is given as:

$$\text{Final Value} = \text{ART} * \text{Victims' Hourly Rate} \quad (1)$$

According to this equation, the avg. Consumer cost becomes $354 for theft resolution. According to 3We, The ART in frauds of new accounts was 26 hours. The average consumer cost for identity theft in the case of a new account was $1205. However, another study by Javelin revealed that the credit card fraud ratio fell by 5.66% from 2017 to 2018 due to the combined efforts of banks, merchants, financial institutes, and card networks [11].

Through these multiple research comparisons, it can be observed that the exact price for the data cannot be determined. However, a vague idea is obtained depending upon the nature of the data. The value for the data varies with the context of the use of data.

VIII. TECHNICAL SOLUTION

There is a dire need for technical solutions to deal with data privacy, accuracy, and ownership issues. Unfortunately, there is no direct way of safeguarding the data. Instead, there are numerous interrelated chunks that provide privacy in certain areas of data, acting as a block of the privacy puzzle. The solution has four focal points.

Firstly, the user sometimes needs anonymity. In order to have it, instead of an online purchase, the user should go to the market to pay by cash. In this way, the user can avoid sharing his data unnecessarily. This may cause a hindrance in buying some specific items, but his credit card data will be prevented from unnecessary sharing.

Secondly, a technology is needed that enables the data to self-destruct. The user should have control over his data. The data has a time limit, after which it is moved to an unknown database and stays there forever. There must be some technology that destroys the data after some time. The research in this field has already been started.

Thirdly, there should be some technology that freezes the signals and data transactions if the data is being collected or stored without the user's permission.

Lastly, there should be some technology that enables the user to track and monitor his data. Out of all four solutions, this one is the most needed. The user must have the information regarding all the data about him being stored and w2here that data is being used.

Though, it is a bit too late to intervene in the data cycle now. However, legal frameworks should be designed, and policies should be changed so that the technology is bound to follow the policies and implement the legal frameworks.

IX. CONCLUSION

Data privacy and data ownership are the two aspects of the identity of an individual. Both the aspects are closely interlinked and are used side by side. The time is near when users will have minimal to no rights to their online data. There are no strong laws or legal frameworks to safeguard user's data, privacy, and identity. The legal community in Australia is working on a proposal that makes recording and monitoring user's activity, private conversations, and recording their data without their permission, a legal offense. The European Union has data protection laws and privacy directives that are implemented in the member states. However, the US has different federal laws for different states. The most major concern is the lack of transparency about where the personal data is being used. As already discussed, data has value. The user whom the data is about must be one of the recipients of that value. Realizing the accurate value may not be possible, but there can be considerations that can help.

Both the govt. And the individuals must take steps to ensure that the ownership of the data remains with the individual whom the data is about. The user must be able to control, monitor, and track his data. Strict laws must be made and enforced regarding data authority, privacy breaches, and non-consensual use of data.

REFERENCES

[1] E. Clarke.,The Underground Hacking Economy is Alive and Well, Secure Works, 18 11 2013. [Online]. Available: https://www.secureworks.com/blog/the-underground-hacking-economy-is-alive-and-well. [Accessed 13 06 2021].
[2] Identity Fraud Survey Report: Consumer Version, (2009).
[3] S.-A. Elvy., Paying for Privacy and the Personal Data Economy, Columbia Law Review, 117(6) (2017) 1369-1459.
[4] C. Gate and P. Matthews., Data Is the New Currency, (2014).
[5] D. Zax, Is Personal Data New Currency? Technology Review, November 2011. [Online]. Available: https://www.technologyreview.com/2011/11/30/20993/is-personal-data-the-new-currency/. [Accessed 10 June 2021].
[6] C. E. Shannon, A Mathematical Theory of Communication, University of Illinois Press, (1949).
[7] W. D. Eggers, R. Hamill, and A. Ali, Data as the New Currency - Government's Role in Facilitating the Exchange, Deloitte Review, (2013).

[8] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, and W. Zhao, A Survey on Big Data Market: Pricing, IEEE Access, 6 (2018).
[9] T. Haselton., Facebook spent $23.4 million in 2019 on Mark Zuckerberg's security and private air travel, 10 April 2020. [Online]. Available: https://www.cnbc.com/2020/04/10/facebook-filing-shows-zuckerberg-and-sandbergs-compensation.html. [Accessed 20 June 2021].
[10] Facts + Statistics: Identity theft and cybercrime, 2020. [Online]. Available: https://www.iii.org/fact-statistic/facts-statistics-identity-theft-and-cybercrime. [Accessed 10 June 2021].
[11] K. Marchini and A. Pascual, 2019 Identity Fraud Study: Fraudsters Seek New Targets, and Victims Bear the Brunt, 2019.
[12] E. Bekker, 2020 Data Breaches The Most Significant Breaches of the Year, Identity Force, 3 January 2020. [Online]. Available: https://www.identityforce.com/blog/2020-data-breaches. [Accessed 10 June 2021].
[13] D. Parent, 500K Zoom Accounts Discovered for Sale on the Dark Web, Fighting Identity Crimes, 15 April 2020. [Online]. Available: https://www.fightingidentitycrimes.com/500-000-zoom-accounts-discovered-on-dark-web/. [Accessed 22 June 2021].
[14] C. R. Team, Family History Search Software Leaks Users' Private Data, WizCase, 20 July 2020. [Online]. Available: https://www.wizcase.com/blog/mackiev-leak-research/. [Accessed 22 June 2021].
[15] B. Heater, Zebra Technologies is acquiring warehouse robotics company, Fetch, Tech Crunch, 01 July 2021. [Online]. Available: https://techcrunch.com/2021/07/01/zebra-technologies-is-acquiring-warehouse-robotics-company-fetch/. [Accessed 01 July 2021].