IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114127

On the Intersection of Big Data and Privacy

Shraddha S. Ghadge, Sheetal A. Wadhai

Second Year Student of Computer Engineering, Universal College of Engineering, Sasewadi Pune

Abstract: A struggle has emerged in relation to the sacredness of one's private information and the importance of moving forward in a digital world of social media, smart devices, and Big Data – an era known as the Age of Context. The purpose of this paper is to make a clear case for concern regarding the seriousness of keeping data private while facilitating efforts to encourage and support emerging technologies. This investigative process included the pursuit of relevant articles and published works that provided a clear and relevant picture of the current state of affairs concerning Big Data and privacy. After a review of the literature, an analysis of data collection methods, a discovery of Big Data processes and purposes, and the identification of risks pertaining to the individual, the military, and the country, it was determined that significant concerns do exist pertaining to data collection, Big Data, and privacy. These concerns not only pertain to the individual, but with military effectiveness and national security.

Keywords: Big Data , Data Security, Data Privacy

INTRODUCTION: -

During the last few decades, the astute observer has witnessed a migration of society towards what has been identified as the Age of Context. This new era is fueled by the engine of social media, which is a blend of applications integrated with mobile phones, portable devices, and smart devices, most of which are operating within the context and framework of the Internet. Applications

The triangulation of the information collected provides a significant volume of data about an individual. This data can be used to describe a person's habits, shopping preferences, health conditions, driving abilities, temperature preferences, likes, dislikes, and a significant host of other self-identifying descriptive variables. Depending on the volume of information posted via social media tools such as Facebook and LinkedIn, collected data will often include self-photos, picture of friends, food preferences, hobbies and outings. In the past, this information was available only to the closest of friends and families; however, social media has made such information available to a much larger audience – some familiar and some not so familiar. In addition, organizations collecting this very personal and private information are capitalizing on its availability and using it for their own self-serving objectives.

Include Facebook, Twitter, Instagram, LinkedIn, YouTube, and other environments utilized to facilitate the social interaction between individuals. It also includes components integrated with television sets, household thermostats, cars, refrigerators, and a host of other smart devices. As noted in [1], social media combined with mobile devices, various sensors, data, and location-based technologies provide a bases for understanding who a person is, what that person is doing, and what that person might do next. The wealth of disparate data and information collected via this complex network of components is a form of Big Data.

Social media constitutes the phenomena of communications processed and stored electronically involving one or more individuals. Merriam-Webster defines social media as electronic communication such as chat rooms, social networking websites, and microblogging by which social communities are created online for the purpose of sharing information, concepts and ideas, and personal communication. [2]. Unlike speaking with a friend in the school hallway or sharing a Polaroid picture at a party, the digital communications garnered from social media is retrieved and stored in databases for an indefinite period of time, unless otherwise altered. Other methods of data collection in today's digital world include smart devices, application usage, customer data collection, and all other digital methods used for the collection and aggregation of information about a specific person, group, or organization. Unsuspecting users may believe their data to be private –for their eyes only; however those parties collecting and managing personal data have a vested interest in benefiting from the data they oversee. Consequently, these data custodians often access data for their own specific purposes, goals, and agendas. The remainder of this paper is organized as follows. In Section 2, Data Collection, we discuss the various methods and processes organizations use to obtain customer and client data. We then proceed with a discussion of Big Data in Section 3 including what it is as well as how it is used. A link is established between Big Data and data collection. In Section 4, risks are highlighted as they relate to the individual, the military, and the country. Finally, we close with concluding remarks.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114127

DATA COLLECTION: -

Companies that facilitate the collection and storage of data often offer this service to the customer's benefit, but more often than not, there are other motives involved. Consequently, data collection might be more appropriately classified as data surveillance. As pointed out by Google, it has seven products that each have at least one billion active monthly users, and that such products could not work as well without access to users' data [3]. This data is often retrieved and used without the customer's knowledge or consent, but Google is only one of many companies to take advantage of the data at their disposal. There are many organizations that collect data unbeknownst to the user. A partial list of applications of both private and government entities that collect data about individuals and use it for their own purposes includes

Automatic License Plate Readers (ALPRs) – INDIAN law enforcement,

• Cookies – Most organizations with websites,

• Facebook tagging system – Facebook,

• Google location services – Google,

• PRISM – NASA,

• Secure flight program – INDIAN Transportation Security Agency (TSA),

• Smart TV's – Various TV manufacturers,

As observed from this list, most of the organizations appear to be reputable entities; however, in the wrong hands, loss of collected data could have serious ramifications for the person for which the data represents. This problem of unauthorized data usage is not just a problem in the US but extends to other places in the world as well. In 2016, Facebook was accused of a breach of Germany's national data protection law through data harvesting using WhatsApp's users, which is its child company [5]. While changes were applied to WhatsApp's terms and conditions, the concern was that those changes were misleading to the consumer as end users were potentially unaware that data collected was being shared beyond the WhatsApp organization. Prior to the last few years, users were oblivious about potential abuse or utilization of their personal data; however, cases similar to the Facebook WhatsApp issue are starting to distill awareness within the abused. In 2017, Facebook was under scrutiny by the French government who accused Facebook of breaching the French Data Protection Act [6]. Again, the issue was related to data sharing between WhatsApp and its parent company Facebook. An Opt-out checkbox was made available to users; however, WhatsApp users had to review the current terms and conditions and then uncheck a box that gave permission to share mobile phone numbers. The phone number was useful to Facebook for the purpose of ad targeting. These incidents are evidence that companies are not transparent about their data usage and sharing activities. This lack of transparency is potentially due to fallout from customers should they understand that their data is being used beyond the advertised purpose. It might also be that some companies are simply defiant when it comes to data they believe to be their own, even while it arguably belongs to the customer.

BIG DATA:-

The term Big Data represents the collection, management, and analysis of both structured and unstructured data. Prior to the introduction of the concept of Big Data, the bulk of data collected by organizations was stored in well-defined data repositories known as Relational Database Management Systems (RDBMS). Structured data represents information collected about payroll records, inventory adjustments, and purchase orders to name a few. With structured data, each field, such as the employee name, birthdate, and pay rate, are defined to be in a specific order, of a specific data type, and of a specific size. In addition, the volume of data is fairly manageable in terms of volume and processing. Unstructured data is less organized and includes such content as video streams, audio recordings, and other data that is not so easy to interpret nor search for regarding a particular content and it is variable by nature. Big Data is defined as a collection of data from traditional and digital sources inside and outside your company that represents a source for ongoing discovery and analysis [7]. Big Data is what facilitates or extrapolates much of the power of the Internet by being able to handle the extreme volume of varied data that traverses the "Information Super Highway".

By being able to manage and search streams of data, regardless of structure, organizations can quickly identify and glean specific bits of data from disparate types and aggregate and summarize it into meaningful information. There are four general reasons in regards to why companies collect and analyze customer data which are as follows

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114127

- Improving customer experience,0
- Refining marketing strategy,
- Turning data into cash flow, and
- Using data to secure data.

Improving customer experience is arguably the primary purpose of collecting and managing data. By improving experience, customers will be happier and will enjoy a richer experience. Companies like Amazon use purchasing related data in order to better assist customers by offering a more tailored environment that is related to the specific customer's interest based on past order history. Refining marketing strategy contributes to this thought of aligning to customer's interest but is far more focused on the goals of the organization to sell more product. Usage information can be used to help pinpoint marketing strategies and thus boost sales. This leads to the third reason for collecting and analyzing customer data, which is turning data into cash flow. Sales will increase while dollars spent on marketing will decrease, resulting in enhanced net profit for the organization.

The last reason noted for collecting customer data is using data to secure data. This is especially relevant in the banking industry where it is important to recognize customer habits and trends in order to prevent misuse or theft. If the organization can identify the behavior of a client, then deviations of behavior can be flagged as a compromise of the account and steps can be taken to lock the account down until the change in behavior and the out-of-the-norm activities can be deemed either authentic or criminal. For example, some banks will allow their users to set up travel notices for their accounts in order to identify when they are going on a trip that will take them out of the area. Without the travel notice, a bank card may be disabled if card use is discovered to be outside the common recorded behavior of the customer. These common practices and behaviors are collected and monitored over time and are considered beneficial to both the customer and the organization regarding the security of the assets of both parties.

RISKS:-

The advantages of collecting structured and unstructured data are many as noted previously; however, their benefits are overshadowed with inherent risks. These risks come in a variety of shapes and sizes and depend on the vulnerability involved and the potential damaged imposed should one be realized. The following narrative highlights various risks as they pertain to the individual, the military, and the nation

RISKS TO A MILITRY:-

While most consider privacy as pertaining to the protection of individuals, it actually extends beyond the bounds of private citizens to include the subject of national security. The military is charged with the protection of its citizens. Its compromise can have detrimental effects on the execution of its duties. In 2018, one such comprise was discovered as it related to the whereabouts of secret military bases due to Heat Map published by Strava, a social network formulated for athletes . The announcement highlighted an update which boasted the display of one billion exercise related activities. These activities were tracked via wearable fitness trackers such as the popular Fitbit. Unfortunately, due to the fact that many of the nation's servicemen and servicewomen are a part of this network, their identities and potential locations of residence became at risk. Considering some soldiers work at secret military bases, tracking their whereabouts indirectly placed the identification of these bases at risk of discovery. This concern was actually realized after an Australian National University student posted his findings on Twitterl

RISKS TO A NATION:-

As previously alluded to, there is a balance between national security and data privacy. This is especially relevant considering threats of terrorism where the enemy lies hidden within the nation's own boundaries rather than presenting themselves as an external aggressor. As a result of successful terrorist attacks such as 9/11, citizens are willing to give up some privacy in exchange for heightened protection against future attacks. In 2018, a poll conducted by the Chicago Council on Global Affairs noted that only one in three people in the United States believe that restrictions regarding data collected by the NSA should be increased [12]. The other two-thirds of those polled believed that current data collection activities should continue. The poll also revealed little support for Snowden, the NSA contractor turned

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022
DOI: 10.17148/IJARCCE.2022.114127

whistleblower. While this was good news for the NSA and its efforts to protect America from terrorism, it begs the question as to how much is enough and at what point should surveillance stop.

CONCLUSION:-

In light of the discussion presenting the pros and cons of the collection of personal and private data by both government and non-government entities, it is evident that concerns are relevant on both sides of the fence. Indeed, it is important to keep private information private and it is equally important, if done in an ethical manner, to allow certain entities access to this data for the benefit of the user of whom the data belongs, the benefit of the organization, and the benefit of the nation. Big Data is the key to housing this information and as such is growing at a phenomenal rate. The basis as to what data an organization can collect, store, have access to, and share is continuing to be litigated in the courts and legislated by state and national law makers. As described in [13], the use of data collection processes and devices and other technologies, and the storage of the information that they capture is being contested in a wide range of course setting, which makes the future use of these technologies likely to be as much a legal issue as it is a technology issue. As we move forward in this exciting Age of Context, social media, and Big Data, let us pursue an ever greater means of integrating our environment without compromising or sacrificing one of our most precious assets, our personal and private information.

REFERENCES:-

1. O'Leary, D. E. (2015). Big Data and Privacy: Emerging Issues. IEEE Intelligent Systems, 30(6), 92--96. https://doi.org/10.1109/MIS.2015.110
2. Verton, D. (2018, August 28). Poll shows Americans more concerned about terrorism than NSA surveillance. Defense. Retrieved from https://www.fedscoop.com
3. Rinehart, W. (2016, June 1). What exactly constitutes a privacy harm? Insight. Retrieved from https://www.americanactionforum.org/insight
4. Popken, B. (2018, May 10). Google sells the future, powered by your personal data. Tech & Media. Retrieved from https://www.nbcnews.com/tech/tech-news
5. Merriam-Webster. (Ed.) (2017) Merriam-Webster Dictionary. Merriam-Webster.com
6. Reynolds, G. W. (2018). Ethics in Information Technology (6th ed.). Boston: Cengage Learning
7. Arthur, L. (2013, August 15). What is Big Data. Retrieved from https://www.forbes.com
8. Kaminski, M. E. (2017). Standing after Snowden: Lessons on privacy harm from national security surveillance litigation. DePaul Law Review, 66, 413--438. https://scholar.law.colorado.edu/articles/724