Security Challenges Faced Due to Increasing Data

Susmita B *, Darakhshinda Parween, V. Nikitha Reddy ****

*, Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science
** Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science

DOI: 10.29322/IJSRP.8.12.2018.p8431
http://dx.doi.org/10.29322/IJSRP.8.12.2018.p8431

Abstract- At the moment Data is one of the most important assets for companies in every sector. The constant demand/growth in the importance and volume of data has resulted in a new problem: it cannot be solved by traditional analysis techniques. This problem was, therefore, handled through the invention of a new paradigm: Big Data. However, Big Data itself came up with new issues related not only to the volume or the variety of the data, but also to data security and privacy of each individual. Therefore, this paper will provide a survey on different privacy and security issues the big data is currently facing. The issues include privacy issues, management issues, integrity security, availability and confidentiality of data.

Index Terms- Security challenges, Infrastructure security, Data privacy, Data Privacy, Integrity and Reactive security.

I. INTRODUCTION

Majority of the social media users share a massive amount of their private information in their respective social network portals. This information include their geographical information, contact details, travel information, images, videos, etc. Many users publish their data publicly without security concerns in their mind. Therefore, social networks have become a large sector of personal data that could be used for any illegal activities. Adding to this, most of the social media users usually have a high level of trust towards the other users on same platform. They accept anonymous friend requests with no security measures taken, and trust any relevant /irrelevant information their friends send them.

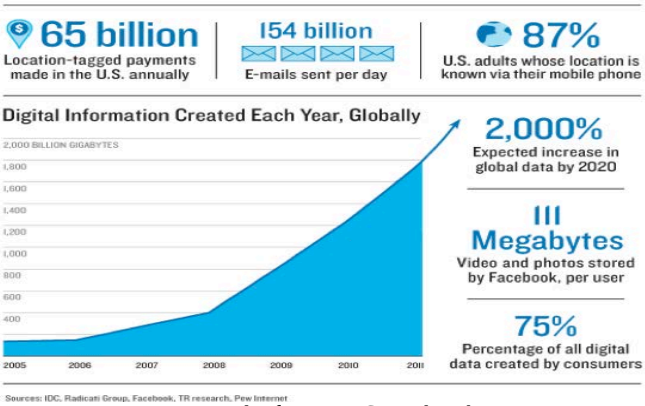

Figure 1: Digital Information Created Each Year

Over the last decade, data privacy has gained importance all over the globe. After the initial whooping use of social networks with little or no attention paid to the lack of control surrounding access to and the disclosure of information, users are now continuously demanding that their data be secured.

Due to social networks large count of users and information available there, and its simple algorithms, social networks have become new platform that urge cyber criminals. This has put all users under great risk. The next part of the paper will talk about variety of privacy and security issues in big data. The issues include privacy issues, identity theft issues, management issues, integrity security, availability and confidentiality of data.

II. LITERATURE REVIEW

The tendency towards increasing the volume and detail of the data that is collected by companies in any sector will not change in the near future , as the rise of social networks, multimedia, and the Internet of Things (IoT) is producing an massive amount of data. We are living in the era of Big Data. Furthermore, this data is mostly unstructured, which signifies that traditional analyzing techniques are not capable of analyzing it. Companies are working to extract more beneficial information from this high volume and variety of data. A new analysis paradigm with which to analyze and better understand this data was needed, hence, emerged Big Data. It was introduced in order to obtain not only private, but also public benefits. But in the process of extracting more and more beneficial data, Big Data have also originated security challenges.

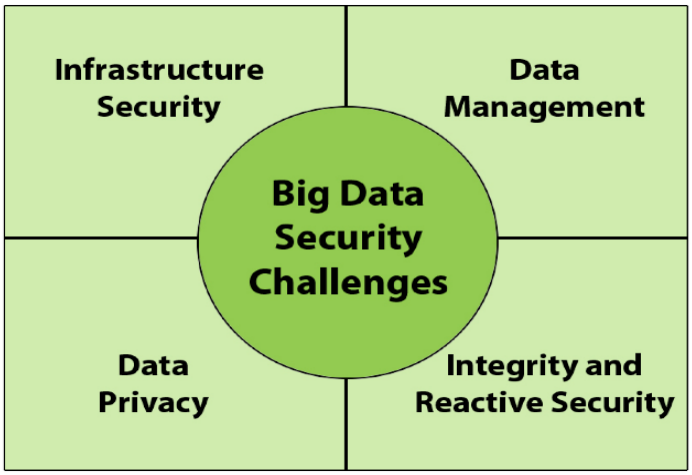

Figure 2: Big Data Security Challenges

1. INFRASTRUCTURE SECURITY

Infrastructure is the cornerstone of Big Data architecture. Owning the right tools for storing, processing and analyzing your data is essential in any Big Data project. Physical infrastructure enables everything and security infrastructure secures all the elements in your big data environment.

2.

As many have previously stated that their Big Data projects handled by them include structured data sources, it is understood that many are using databases containing sensitive/personal information (such as financial data, marital data, academic data etc.) which are still available on-site. They are server hugging. They do not want to loose control due to security issues, fear of implications if data was lost, deleted or accessed. Keeping the infrastructure on-premise gives a sense of security.

Figure.3: Evolving Threats to Data Infrastructure

2. DATA PRIVACY

Data privacy is probably the issue about which the social networking users are most concerned, but it should also be one of the biggest concerns for the organizations that use Big Data techniques. A Big Data system usually contains a massive amount of sensitive information that organizations use in order to make the data benefactor for them. However, we are unaware of where the limit regarding the use of that information is. Organizations should not be given total freedom to use that information without our knowledge/permission, although they also need to gain some benefit from that data. Several techniques and mechanisms with which to secure the privacy of the information and also allow organizations to still be benefited from it have therefore been developed, and attempt to solve this problem in various different ways.

The current will be discussing two privacy issues:

1) Users' Anonymity
2) User's Profile and Personal Information

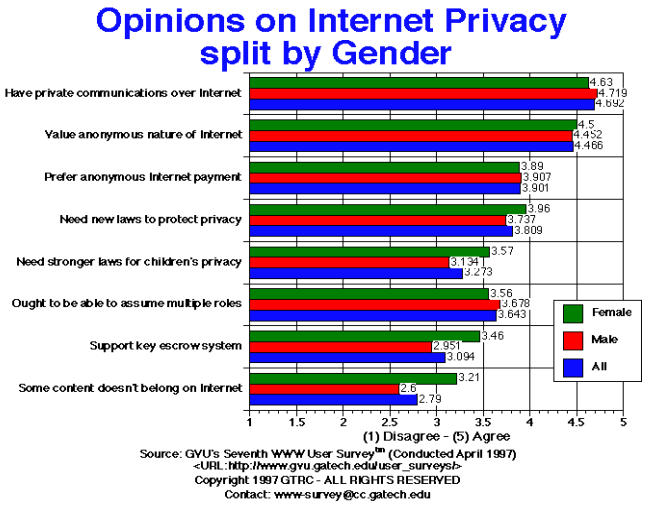

Figure 4: Internet Privacy

2.1 USERS' ANONYM

In almost all the social networking websites, users tend to use their original name to create their accounts. So, their identity is exposed publicly to peer users and any illegal users, as well as all the anonymous users in the online world. Moreover, all the social network user's account is being indexed by many search engines such as Google, Bing etc., and automatically flashes in the top positions for results of the search. This goes in favor of the attackers, if he/she possess the knowledge of names of the victims, they can happily search for victim's profile and obtain the data he/she requires, or they can search through social networking sites to target new preys. Keeping aside the use of original name as account name, there are also other methods that are utilized to invade social network user's anonymity. The two methods are de-anonymize attack and neighborhood attack. De-Anonymization Attack was published by Gilbert Wondracek and his team proved that by possessing and utilizing information of the members in a group and the technique of stealing the history, villain can easily decrypt anonymity of their prey [5]. In this method, attackers should have knowledge regarding which group on social network (number of people/users that possess similar interests or number of people that come from same background e.g. attended the same college or have same hobbies) preys belong to. A group on social networking site is being targeted as the total groups on social media is less than the number of individual users, which makes it easier to focus on the group, and then that group would help the attackers focus an individual user. Attackers will make use of history-stealing method to gain the information regarding URLs (websites) that preys visited in the past in order to cluster out victim's group. Before explaining about how this technique works, let us discuss about social network link and history stealing. There exist 2 kinds of links in social networks. A static link – It displays the home section of the users, which is pretty similar for all users. Dynamic link – It clearly specifies the uniqueness of every individual or a particular group. Coming to history stealing methods, attackers/hackers drag users to their respective web pages very effectively, and then try to get hold of user's browsing history by simply sending out a particular number of URLs to the

users(probably the one's which the user would use). The group directory which is provided by the social networking sites enables the attackers to get the URLs information. Then, attackers will try and check the user's history in order to know if he/she has visited the URL on the list or not. Later, the browsing history information will be sent back to attackers. "This process of getting the browsing history can be done by making use of conditional logic in CSS (Cascading Style Sheet) i.e. a: visited and display: attribute" or JavaScript (client-side script). Therefore, with the use of history stealing technique, attackers obtain prey's browsing history, and use the browsing history in order to filter out the URLs visited by him/her, here the importance will be given to dynamic links as they possess the unique information about the users. In general, social networks have the information of mailing list of the group members. So, attackers can utilize the acquired emails to get hold of identity (profile) of their preys.

2.2 Neighborhood Attack

Social graph can be used to represent the social networks. In social graph, a social network user is indicated by a node, and relationship between two social network users is indicated by an edge. The process of Neighborhood attack is that if an attacker has the details of the victims' nearest neighbors' node and the relationship between them, they can track their preys' node. Let's take an example; the data of neighbors of the prey is available to the attacker. Data is of five friends of prey A.A friends are B, C, D, E, F. Among them B and C are friends with each other whereas, D, E, and F are not. Figureure 5 can be used to represent neighborhood attack on A. This graph can be used by attackers in order to track A since every social networking node has a unique neighboring attack graph.

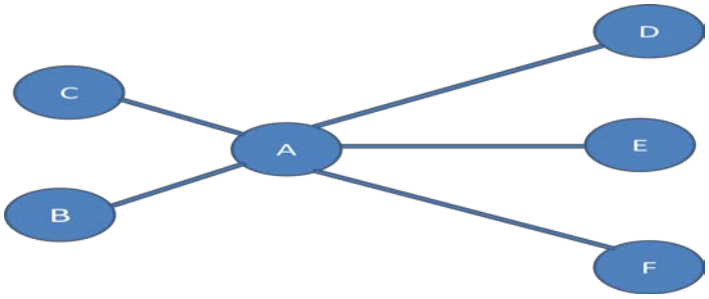

Figure 5: Neighborhood Attack in Social Networks

2.3 User's Profile and Personal Information

Almost all the profiles of users on social networking sites have the original information given by the users. Sensitive information include their geographical information, contact details, travel information, images, videos, date of birth, professional life details and academic details urges attackers to target them. Hence, the major issue of user's account is the misuse of profile and personal details.

What are sources of leakage of users' profile?
Due to lenient privacy policies: Not many users are bothered about the privacy policies of their account. The audience is set as public so anyone can access their data without any prior permission. Also, many social networking land site no remittal concealment exercise set is still not safe such as in Face book, a ally of a ally who the substance abuser does not know can still see his information. However, even the safest secrecy circumstance; there are still flaws that allow attackers to access user's information

Selling out information to 3rd party members: An API (Application Program Interface) is given out to the third party developers by the social networking websites such as Face book in order give them access to develop their applications on the website. Social networking users are very well aware of these third party developers as well. Once the third party applications get the access to the data of the users', they can anytime use the data. It is also capable of publishing any unwanted information on user's page or user's friend's page without user's knowledge. Leakage of information to 3rd party domain: A lot of social networking sites these days are giving away the access to user's information or their activities to the 3rd party domains for the sake of their commercial benefit.

Identity Theft: The usage of one's identity or information as their identity in order to perform any malicious activities is known as identity theft. Attackers usually get attracted to the social networking sites as they tend to have continuous increase in the amount of information shared on them. Profile cloning is one of the types of identity theft. In this technique, people who are not careful are easily trapped and the trust of friends is also taken as an advantage, when they accept friend requests. Social phishing is another method that can be used to steal social network user's identity.

III. DATA MANAGEMENT

Data management is the organization, administration and governance of immense volumes of both structured and unstructured data.

The goal of immensely colossal data management is to ascertain a high caliber of data quality and accessibility for business astuteness and sizably voluminous data analytics applications. Corporations, regime agencies and other organizations employ sizably voluminous data management strategies to avail them contend with expeditious-growing pools of data, typically involving many TiB or even petabytes of information preserved in a variety of file formats. Efficacious sizably voluminous data management avails companies locate valuable information in immensely colossal sets of unstructured data and semi-structured data from a variety of sources, including call detail records, system logs and gregarious media sites.

Figure 6: Data Management

3.1 Profile Cloning

One technique of stealing social network user's identity is called profile cloning. The main targets of profile cloning are users who set their profiles to be public. Public profile sanctions assailers to obtain profile information facilely, and ergo can duplicate or copy their profile information to engender an erroneous identity. There are two types of profile cloning. Subsisting Profile Cloning In subsisting profile cloning, assailers engender a profile of already-subsisting users by utilizing their denomination, personal information, as well as picture to increment reliance, and then sending friend requests to friends of that utilizer. This action is prosperous since most users accept friend requests from the person that they already ken without looking through it conscientiously. Additionally, it is possible that a person might have multiple accounts. If victims accept the friend requests, then assailers will be able to access their information. Cross-Site Profile Cloning In cross-site profile cloning, assailants purloin user's profile from one gregarious networking site that users register an account, and then engender an incipient user's profile on another gregarious networking site that utilizer has not registered on afore. After that, assailants use users contact list from the registered convivial networking site to send a friend requests to all those contacts in another convivial networking site. In this case, it is more convincing than the first case since there is only one account for that particular utilizer. Then, if the contacts accept friend request, assailants can access their profile.

3.2 Social Phishing

In phishing attack, attackers provide a fake website that looks authentic to lure victims into providing their sensitive information such as password, financial information, or identification number to the website. Phishing attack together with personal information from social networks make the attack becomes more successful. Attackers can use the social engineering method by gathering data from social network users and then perform automated extraction of data to obtain context-information that is useful to trick users to the phishing site. For example, attackers can send a phishing website to victims by using the victim's friend's names.

IV. Integrity and Reactive Security

In the end all that a user needs from any social networking site is that the site has to assure the three main aspects to be secure. They are Confidentiality, Integrity and Availability. These three are called as CIA triad, which is designed for information security within an organization.

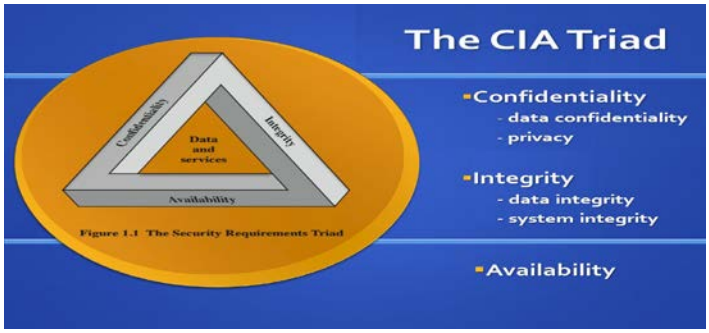

Figure 7: CIA Triad

4.1 Confidentiality

Although privacy is traditionally treated as a part of confidentiality, we decided to change the order owing to the tremendous impact that privacy has on the general public's perception of Big Data technology. The authors that approach this problem often propose new techniques such as computing on masked data (CMD), which improves data confidentiality and integrity by allowing direct computations to be made on masked data, or new schemes, such as Trusted Scheme for Hadoop Cluster (TSHC) which creates a new architecture framework for Hadoop in order to improve the confidentiality and security of the data.

4.2 INTEGRITY

Data integrity is the maintenance of the accuracy and assurance of data over its entire life cycle.

Any unintended changes to data as the result of a storage, retrieval or processing operation, including maleficent intent, unexpected hardware failure, and human error, is failure of data integrity. If the transmutations are the result of unauthorized access, it may withal be a failure of data security. Depending on the data involved this could manifest itself as benign as a single pixel in an image appearing a different color than was pristinely recorded.

4.3 AVAILABILITY

Data availability is the process of ascertaining that data is available to culminate users and applications, when and where they require it. It defines the degree or extent to which data is yarely utilizable along with the indispensable IT and management procedures, implements and technologies required to enable, manage and perpetuate to make data available.

Availability is best ascertained by rigorously maintaining all hardware, performing hardware repairs immediately when needed and maintaining a correctly functioning operating system environment that is free of software conflicts. It's additionally consequential to keep current with all indispensable system upgrades. Providing adequate communication bandwidth and averting the occurrence of bottlenecks are equipollent consequential Redundancy, failover, RAID even high-availability clusters can mitigate earnest consequences when hardware issues do occur. Expeditious and adaptive disaster instauration is essential for the worst case scenarios; that capacity is reliant on the ease of a comprehensive disaster recovery plan (DRP). Safeguards against data loss or interruptions in connections must include capricious events such as natural disasters and fire. To obviate data loss from such occurrences, a backup copy may be stored in a geographically-isolated location, perhaps even in a fireproof, waterproof safe. Extra security equipment or software such as firewalls and proxy servers can sentinel against downtime and unreachable data due to maleficent actions such as denial-of-accommodation (DoS) attacks and network intrusions.

V. CONCLUSION

This paper provides an explanation of the research carried out in order to discover the main problems and challenges related to sec urity in Big Data, and how researchers are dealing with these pro blems. The objective was achieved by following the systematic mapping study methodology, which sanctioned us to find the papers cognate to our main goal. Having done so, we discovered that the principal quandaries are cognate to the innate characteristics of a Astronomically immense Data system, and additionally to the fact that security issues were not contemplated when Astronomically Immense Data was initially conceived. Many authors, ergo, focus their research on engendering betokens to forefend data, categorically with deference to privacy, but privacy it is not the only security quandary that can be found in a Immensely colossal Data system; the traditional architecture itself and how to bulwark a Hadoop system is withal an astronomically immense concern for the researchers. We have, however, additionally detected a lack of investigations in the field of data management, especially with veneration to regime. We are of the considered opinion that this is not acceptable, since having a regime security framework will sanction the rapid spread of Sizably Voluminous Data technology. In conclusion, the immensely colossal Data technology seems to be reaching a mature stage, and that is the reason why there have been a number of studies engendered the last year. However, that does not betoken that it is no longer compulsory to study this paradigm; in fact, the studies engendered from now should fixate on more categorical quandaries. Furthermore, Astronomically Immense Data can be subsidiary as a base for the development of the future technologies that will transmute the world as we optically discern it, like the Internet of Things (IoT), or on demand accommodations, and that is the reason why Astronomically immense Data is, after all, the future.

REFERENCES

[1] Antoine Fressancourt and Maria Pernas Martinez, Ascent / "Privacy and social networks" paper.

[2] Julio Moreno, Manuel A. Serrano and Eduardo Fernández-Medina Alarcos Research Group, University of Castilla-La Mancha, Article on "Main Issues in Big Data Security".

[3] Christopher F. Spinelli Corporate Communications Elon University, "Social Media: No 'Friend' of Personal Privacy paper".

[4] Dolvara Gunatilaka , "A Survey of Privacy and Security Issues in Social Network" paper.

[5] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel, " Practical Attack to De-anonymize Social Network Users," IEEE Symposium on Security and Privacy, 2010, pp.223-238. http://iseclab.org/papers/sonda-TR.pdf .

AUTHORS

Darakhshinda Parween – M.Tech. , Sree Dattha Institute of Engineering and Science, Hyderabad, Telangana , India, dparween2020@gmail.com.
Susmita B – B.Tech. Sree Dattha Institute of Engineering and Science, Hyderabad, Telangana , India, susmitabreddy@gmail.com.
V Nikitha Reddy – B.Tech. Sree Dattha Institute of Engineering and Science, Hyderabad, Telangana , India,

Correspondence Author –Darakhshinda Parween, dparween2020@gmail.com. , 9955886240.