Security Sensitivity in Big Data

[1]Shipra Varshney, [2] Ekta Jain, [3] Nishi Sharma

Department of Computer Applications, Northern India Engineering College, New Delhi

1shipra_vin@yahoo.com

2g.ektajain@gmail.com

3nishisunil17@gmail.com

Abstract - A situation which nowadays has been arising from the accrue number of data collected from several sources, including the internet is known as Big Data. Big data should not limit to size or volume. Big data has certain specific characteristics (volume, variety, velocity, and value - 4V) that make it difficult to handle from security point of view. The advancement and the growth of data to become big data raise another important concern about data security and its management. Many organizations like NIST explains a guide for conducting risk assessments on data, including risk management process and risk assessment. In this paper we have highlighted at various risk management guidance and identifies whether the method of this standard is valid and pertinent to big data by normally define the threat source, fear events, exposure, likelihood of occurrence and effect. The result of this study will be a general structure defining security aspects and management on big data. [1]

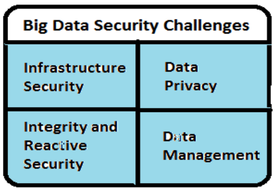
***KEYWORDS* —BIG DATA, NIST, COMPUTATIONAL METHODS, EU DATA DIRECTIVE**

I. INTRODUCTION

THE biggest challenge from a security point of view for big data is to safeguard of user's privacy. Big data often contains large and huge amounts of personal distinguishable information and so the privacy of users is an important concern. For the above said reason, the big amount of data stored violates security and affects in more destructive consequences than the data breaches we usually see in the real world. So, a big data security rupture will potentially affect a large group of people, with the outcome not only from a standing point of view, but with large legal consequences.

The information generation for big data in the organizations, they have to make sure that they have the appropriate equilibrium between the usefulness of the data and privacy. Before the data is saved and stored it should be fairly anonymised, removing any unique identifier for a user. Now this method of removing identifiers might not be enough to guarantee that the data will remain anonymous in itself is a big security challenge. The de-anonymization techniques could be applied and cross referenced on this available anonymous data.

data so that new encrypted data will be created. And on decryption the results will be the same just like if the operations were carried out over plain text data (original one). So, without the knowledge of the underlying plain text data, the cloud will be able to perform operations over encrypted data. [2]



Big Data Security Challenges	
Infrastructure Security	Data Privacy
Integrity and Reactive Security	Data Management

Fig 1. Main challenges in regards of Big Data Security. [4]

II. PRACTICES AND PERSPECTIVE

Big data is absolutely a new concept and therefore security community has not yet implemented its best practices and policies on it. Nearly all data security issues are caused by the shortfall of functional and effective measures provided by antivirus software and firewalls. As Big Data goes beyond hard disks and isolated system but these systems were developed to protect the limited scope of information stored on the hard disk. However there are numeral general security recommendations that can be applied to big data which are listed below:

--First, Evaluate our cloud providers: If we are storing our big data in the cloud, we must safeguard that our provider has suitable and sufficient protection mechanisms in place. Also make sure that periodic security audits should be carried out and it should also agree on the penalties if the standards are not met.

--Second, Create a proper access control policy: Design policies that allow access to authorized users only.

--Third, Secure the data: Both the raw data and the outcome from analytics. It should be adequately protected. To ensure the encryption can also be used accordingly so that no sensitive data is leaked.

--Fourth, Save and protect communications: Data in transit should be appropriately protected to confirm its confidentiality and integrity.

--Fifth, Deploy real-time security monitoring: Access to the data should be observed by Cyber Threat intelligence (CTI) and by using it we can prevent unauthorized access to the data. Actually CTI includes in-depth information about specific threats which organizations can protect it from the types of attacks that could do them the most damage. [3]

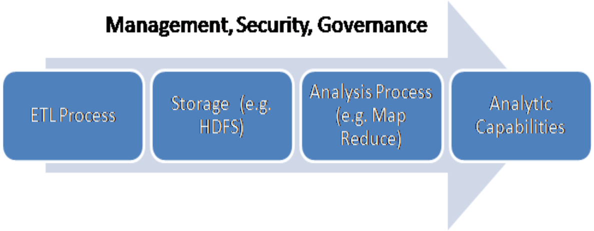
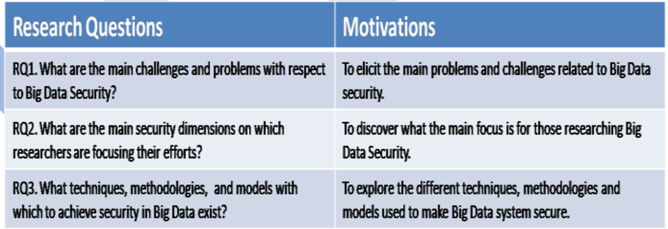

Management, Security, Governance
ETL Process
Storage (e.g. HDFS)
Analysis Process (e.g. Map Reduce)
Analytic Capabilities

Fig 2. Big Data Architecture

I. Main Research Questions

In our case, the questions include the inspection of the common but main challenges and problems that can be found with respect to the topic of Big Data security. Along with another question whose idea is to discover the main security areas and measures on which researchers are focusing their efforts. Finally, we want to discover which different techniques, practices or models have already been developed in order to deal with these problems.

The main Research questions are as follows:
RQ1. What are the main challenges and problems in Big Data security?
Meaning: To generate the main problems and challenges related to Big Data security.
RQ2. What are the main security dimensions on which researchers are focusing their efforts?
Meaning: To uncover what is the main focus for those researching in Big Data security.
RQ3. What techniques, methodologies, and models to achieve with security in Big Data exist?
Meaning: To survey and search the different techniques, methodologies, or models used to make Big Data systems secure.



Research Questions	Motivations
RQ1. What are the main challenges and problems with respect to Big Data Security?	To elicit the main problems and challenges related to Big Data security.
RQ2. What are the main security dimensions on which researchers are focusing their efforts?	To discover what the main focus is for those researching Big Data Security.
RQ3. What techniques, methodologies, and models with which to achieve security in Big Data exist?	To explore the different techniques, methodologies and models used to make Big Data system secure.

Fig 3. Research questions and their motivations.

II. Technological Solutions

The adequate use of encryption is the main solution to ensuring that data remains protected. For example, Attribute-Based Encryption can help in providing subtle access control of the encrypted data.

The data is also important to guarantee that privacy concerns are handled. It should be ensured that all delicate information is removed from the set of records collected.

Real-time security monitoring is also a key security component for a big data project. Organizations should ensure monitoring that there is no unauthorized access. More sophisticated attacks are detected using threat intelligence and that the organizations can react to threats accordingly. [3]

III. Strategic Issues

A mechanism called **risk assessment** over the data should also be run by the organizations over the data they are collecting. It is the job of the organizations to consider whether they are gather any customer information that should be kept private and establish appropriate policies which protect the data and the right to keep the privacy of their clients.

If other organizations also demands the same data means if the data is shared with other organizations then it should be considered how this is done. Purposely released data that turns out to violate on privacy can have a great and massive impact on an organization from an eminence and economic point of view.

Organizations should also cautiously consider regional laws around griping customer data, such as the ***European Union Data Directive*** known as ***EU Data Directive***. This is a regulation adopted by the European Union to protect the privacy and protection of all personal data collected for the citizens of the EU, especially as it relates to processing, using or exchanging such data.

IV. What next in Big data Security?

The concern of the security of the cloud-based systems is of immediate concern to companies using Big Data. Intel Security has recently published the McAfee Labs' Threat Predictions Report that holds their predictions for the near-future of data security. A report with a particular concern is the idea that acknowledges cloud files hosting services such as Drop box, Box, and Stream Nation, is at risk of being used as control servers in upcoming cyber surveillance campaigns. If targeted, these popular cloud services could enable the malware to transfer commands without raising suspicion.

Mischievous attacks on IT systems are becoming more intricate and new malware is continuously being developed. Sadly the companies that work with Big Data face these issues on a daily basis. Still every problem has a solution and finding an effective and suitable answer for your organization is really possible. [3][6]

V. Privacy Preserving

Apparently the most challenging and anxious problem in Big Data is security and privacy. The health care industry, governmental agencies, biomedical researchers, and private businesses devotes enormous resources into the collection, accumulation, and sharing of large amounts of personal data for the stupendous benefit of Big Data. In one of the surveys the National Security Administration routinely collects and examines huge amounts of personal data obtained from heterogeneous data sources such as the Internet, telecommunications and the user databases of large businesses, including Google, You Tube Microsoft, Yahoo, Face book, Pal Talk, AOL, Skype and Apple". The surveys and many facts show that Big Data will damage and harm the user's privacy if it is not properly handled and managed.

The security and privacy issues which should be concerned in Big Data context include:

1. Information concerning the users (people) is collected and used in spick and span to add value to the business of the organization. This is done by creating awareness in their lives which they are unaware of.
2. Another important consequence arising would be Social stratification where a literate person would be taking advantages of the Big data predictive analysis and on the other hand under-privileged will be easily identified and treated worse;
3. Big Data used by law enforcement will increase the chances of certain tagged people to suffer from adverse consequences without the ability to fight back or even having knowledge that they are being discriminated
4. A bunch of challenges in the field of privacy in big data involves interaction with person, re-identification attacks, probable and provable results, and economic effects. Communication with individuals includes providing transparency, obtaining consent, and revocation of consent and deletion of personal data. [4]

Governance frameworks

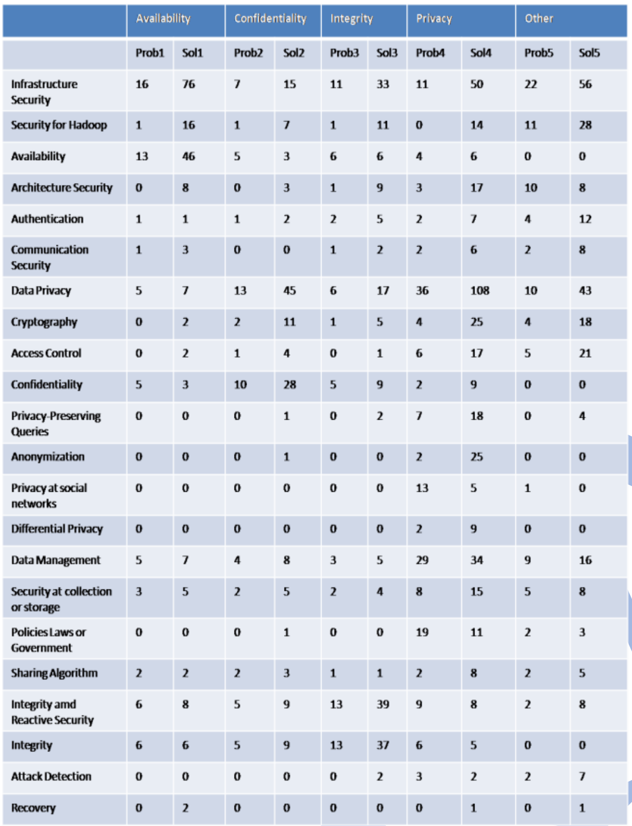
The Data collected for Big Data could be misleading if an adequate governance framework is not used in Big Data and will cause unexpected raise in costs.

Because of the evolution of new concept Big Data, governance framework is not able to create procedures and policies. Data collected so far is from the external sources, which is a challenge as the nature of data is highly unstructured. It makes it difficult to model, categorize and map the data when it is stored and captured. [5]

Organizations have to be very careful in identifying what data is to be collected and will be beneficial to the organization in value addition. If the data is collected and not filtered properly, it will be of high risk wasting time and adding no or little value to the business. [10]

III. RESULTS

We are hoping that this paper will bring a stimulus in mind to enhance the research and development community to focus on the research methodology. The results achieved so far in this area are represented in the below figure. The objective of the below figure is to make it easier to visualize the main problems and their relation to the security dimensions. Each column shows the number of papers found, and there are two outlooks: on the one hand, those papers that explain the problem with respect to one specific topic, and on the other, those that deal with the problem and propose a solution. [7]



	Availability		Confidentiality		Integrity		Privacy		Other	
	Prob1	Sol1	Prob2	Sol2	Prob3	Sol3	Prob4	Sol4	Prob5	Sol5
Infrastructure Security	16	76	7	15	11	33	11	50	22	56
Security for Hadoop	1	16	1	7	1	11	0	14	11	28
Availability	13	46	5	3	6	6	4	6	0	0
Architecture Security	0	8	0	3	1	9	3	17	10	8
Authentication	1	1	1	2	2	5	2	7	4	12
Communication Security	1	3	0	0	1	2	2	6	2	8
Data Privacy	5	7	13	45	6	17	36	108	10	43
Cryptography	0	2	2	11	1	5	4	25	4	18
Access Control	0	2	1	4	0	1	6	17	5	21
Confidentiality	5	3	10	28	5	9	2	9	0	0
Privacy-Preserving Queries	0	0	0	1	0	2	7	18	0	4
Anonymization	0	0	0	1	0	0	2	25	0	0
Privacy at social networks	0	0	0	0	0	0	13	5	1	0
Differential Privacy	0	0	0	0	0	0	2	9	0	0
Data Management	5	7	4	8	3	5	29	34	9	16
Security at collection or storage	3	5	2	5	2	4	8	15	5	8
Policies Laws or Government	0	0	0	1	0	0	19	11	2	3
Sharing Algorithm	2	2	2	3	1	1	2	8	2	5
Integrity amd Reactive Security	6	8	5	9	13	39	9	8	2	8
Integrity	6	6	5	9	13	37	6	5	0	0
Attack Detection	0	0	0	0	0	2	3	2	2	7
Recovery	0	2	0	0	0	0	0	1	0	1

Fig 4. Collective Results.

IV. CONCLUSION

This paper provides an insight of what research is carried out to solve the main issues and challenges related to security and privacy in Big Data, and how the researchers are proceeding with these problems. This manifest was attained by following the meticulous mapping study methodology, which helped us to search for the subjective papers related to our goal. After completing this we have found that the primary problems are related to the inherent features of our topic Big Data system, and the fact that security issues were not examined when Big Data was initially originated. Many authors, therefore, shown their concern on research on how to protect data, specially with respect to privacy, but there are other problems also in the system Big Data; the architecture itself given originally and the System which supports Big Data a Hadoop is also a big challenge for the researchers. In conclusion, the topic Big Data technology is now reaching a mature stage, and that is why large number of studies was created last year. Still, we have found that the there is lot of scope of improvements; we should focus more on the specific problems rather than studying the same paradigm. Furthermore, this topic can be useful as a base for handling the upcoming challenges, be it development or future technologies for the development of the future technologies that will bring the positive change in the world as we have seen it, like on demand services and the Internet of Things (IoT), or, and that is how there is lot of scope in Big Data.

REFERENCES

[1] Expanded Top Ten Big Data Security and Privacy Challenges (Book style with paper title and editor)," by Big Data Working Group.

[2] Big Data Survey Results Show that Data Security is Paramount By Cynthia Leonard — November 5, 2015

[3] Big Data Security - Challenges & Solutions by Guillermo Lafuente, 10 November 2014.

[4] A Survey Paper on Security Issue with Big Data on Association Rule Mining by Prof. Asha Patel Assistant Professor Department of Computer Engineering SAL Institute of Technology & Engineering Research ,Gujarat, India

[5] Kalyani Shirudkar, Dilip Motwani Big-Data Security International Journal of Advanced Research in Computer Science and Software Engineering

[6] Raghav Toshniwal* Kanishka Ghosh Dastidar Asoke Nath Big Data Security Issues and Challenges International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 2, Volume 2 (February 2015)

[7] https://www.researchgate.net

[8] http://www.datacenterknowledge.com/

[9] Security and Privacy –A Big Concern in Big Data A Case Study on Tracking and Monitoring System (IJIRST/Conf/NCLTNCS/2017/007)

[10] Big Data Security Issues, International Journal of Technical Research & Applications e -ISSN: 2320-8163, www.ijtra.com, Special Issue 42 (AMBALIKA) (March2017), PP. 21 25