

The Discovery for Privacy & Security in Various Big Data Application: A study

Prince Vivek, Rajeev Ranjan

School of CSA, REVA University

Rukmini Knowledge Park Yelahanka, Kattigenahalli, Bengaluru, Karnataka 560064

Jaiswalvivek005@gmail.com

rajeevranjan@reva.edu.in

ABSTRACT

This paper is aiming to provide a literature re-examine on the need of data safety and privacy issues of various big data applications. In this, the first section gives a brief description of big data. The second section reviews the application of big data and, hence, explains the importance of privacy and security of Big Data in the third section.

Index Terms: Big Data, Application, Security, Privacy

I. INTRODUCTION

As per the researchers of Big data, Big data is described as the large volume of data, Fig. 1: shows the popularly known 5Vs (for data authenticities we have emphasize and made 5 from initial 3 V's) they are as fallows volume, velocity, variety, veracity and value, as shown in

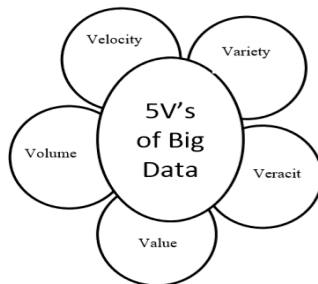


Fig. 1 Five V's of Big data

Volume refers to the amount of data generated and measured in terabytes (240) or pet bytes (250), and is rapidly heading toward Exabyte's (260). The rate at which new data are generated is often characterized as velocity, and data production occurs at very high rates. Data are heterogeneous and can be highly structured, semi-structured, or totally unstructured, and is denoted by variety. Veracity is due to the diversity among data sources, data evolution and raises concerns about security, privacy, trust, accountability, and creating a need to verify secure data provenance; and value measures the usefulness of data in making decisions [1, 2]. Traditional data processing applications are not sufficient to process big data with so large or complex data sets. Current years the amount of data generated by internet, social networking sites, sensor networks, healthcare applications, vast research works and organizations, is drastically increasing day by day. Dealing with Big Data includes capture, storage, dissemination, search, analytics and visualization when typical problems encountered. A major challenge for senior executives, IT researchers and practitioners due to the exponential growth of data are: (1) design appropriate systems to handle the data securely to provide privacy and

(2) analyze it to extract the intended data only to provide relevant meaning for decision making to the targeted people.

II. BIG DATA APPLICATIONS

Some of the big data applications are batch processing types and requires complex calculations and multiple iterations over entire data. Some of them are real time in nature which needs low latency requirements and process incoming stream of data.

A. Social Media

Social media, internet-based media, helps the individuals to interact with one another, exchanging personal details such as biographical data, professional information, personal photos, texts, messages and instant thoughts, question papers, assignments, and workshops in the Education domain, online survey, marketing, and targeting customers in Business land, and jokes, music, and videos in Entertainment domain. Most of the social media started to interact with friends and relatives and for entertainment, but now it is used in businesses to reach out to customers, for marketing their products. Moreover, the amount of photos uploaded to the social media is growing rapidly and a majority of these photos have no privacy relevance. Data reside in more dispersed form and it is bridging users of a larger and heterogeneous network.

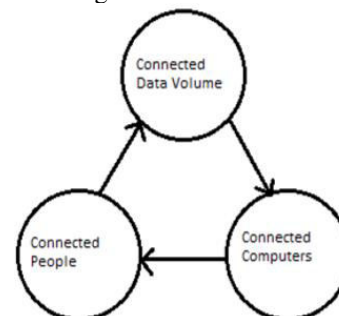


Fig. 2 People in the network produce a data stream

The amazing growth and diversity in connected data continues to profoundly affect how people make sense of this data. Connected people in the network produce an incessant data stream that is deposited into a repository of connected data. Big data analytics should be done on these data by clouds or connected computers; and these

computers generate intelligence that subsequently proliferates back to connected people [3]. In rising time of Big data, applications were majorly working at non personal data. But today majority of big data applications resides in social domains such as social web applications, consumer analytics and governmental analysis.

B. Modern Industry

Sensors embedded in modern machines produce 1000 Exabyte of data annually and is expected to increase many fold in the next ten years [6]. The bad decision which is taken by the organizations disrupts the customers and everyone in between. Big data helps the management to take best decisions by the number of correlations and statistics. This updated system providing a way to the people to the correctly value their data, privacy and information security. This would be a major step in the field of financially feasible private and secure alternatives. This lead to an age where people can maintain their information privacy and retain ownership and control over their digital assets they produce and maintain. The potential benefits and challenges of big data will naturally differ from sector to sector. Various areas in public and private sectors such as product and market development, operational order, market demand prophecy, decision making, and customer experience and loyalty are benefited by the use of big data. It is seen from Fig. 3 Customer behaviour can be analysed with the help of data retrieved from sensors embedded in smart products.



Fig3: The use of big data by the respondent

Preventive measures can be ensured prior to the failure of the product. This way, big data can be used to refine the development of the next generation of products and services. Vulnerability of big data is much higher because of its huge size, broader distribution over the network and broad range of access by different means. In addition, many softwares do not take security as seriously enough as it should be taken, this also goes with the company infrastructure. This may lead to security threads. as we know big data deals with large distributed network. Majority companies have been using Hadoop as their tool. Even though it's popular and has access controlled

mechanism it is wasn't created for such a large scale data to be handled.

Industrial Big Data in internet:

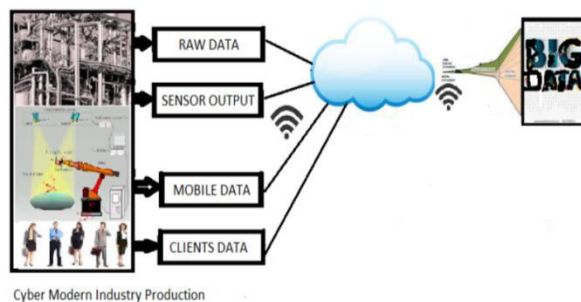


Fig 4.Factory Big Data connected with the internet

In the factories, the machines and resources of an enterprise is connected to the web services directly to the customer or by outer parties. The machine tools are enhanced with Information that is given by the analysing big data. Business management instruments will be attached to the machines to capture various data from different sources. A main challenge towards the transformation to cyber-physical production systems is to propose of standards and safe communication protocols which should be capable of interfacing existing systems and to collect and exchange information and processed data.

Finance

The financial analyst wants to correlate company's proprietary data with data on the network storage to know about some kinds of investment or business opportunities. He also wants to correlate public and private data to study about market trends and new opportunities. These types of network storage access leads to security vulnerabilities such as man-in-the-middle attacks. The sensitive information should be intercepted or disrupted due to its high value and huge traffic. The usual procedure to protect data using firewall or to guard it by providing information security such as encryption or hashing techniques cannot be used for big data due to its size and distributive organization. The need for new cyber security tools or frameworks to be developed to prevent the emerging security vulnerabilities in financial networks. These tools or framework should support big data analytics in real-time networks supporting rapid trading, credit card verification, and other financial services related to consumer banking, investment banking, and their supporting data infrastructures. The essential data sharing between companies can be encouraged by protecting proprietary information. Attackers are searching for new methods to accomplish financial gain. The money is stolen not only by selling the stolen data but also through sabotage and fabrication of data records or transactions. Financial/banking industry has to make sure it is on top of such threat path. IBM's X-Force [10].

Transportation

Real-time big data analytics is very much needed in Transportation system which provides various kinds of information to travellers within a very short time. Real

time applications guarantee "deadlines" such that it requires all the needed resources available while processing. Real-time big data analytics in transportation helps vehicle tracking, route selection for the destination, estimate time to reach to the destination, change route due to some incidents, find the fastest route for emergency delivery of items and dynamic route identification for emergency vehicles for the quick arrival to the destination. Sensor technology is used for Monitoring Traffic conditions and is connected to the communication technology.

Stock market

A pool of share buyers and share sellers for registered companies is what a stock market is. There is a very huge data generation in this domain. And the data generated is also very versatile in nature. As this is a business centre there are high possibilities of frauds being taken place. Big data analysis could solve this by detecting the frauds or any illegal activities happening in real time. This will improve the overall performance of the stock market. Many softwares companies are actively working in share predictive analysis by taking advantage of the huge dataset available in the market. Software could predict the up and downs of any share value which can lead to high margin profit. This wasn't the case in earlier days. People used to rely on the stock broker and only on his experience in the market. We can make out that stock brokers experience is not relabel as it varies and can be different many times there could be no logical proof for that. But a machine learning software will give always logical conclusion about any stock price based on the history. Taking a decision based on the Bigdata analysis prediction could turn in to a calculative and logical risk. And ofcourse there are many parameters which machine may not take into consideration such as emotions of the public towards the company or its product. But that's a different argument. These improvements can bring in more people to invest in the market which can increase the participation and in overall market could grow and operate in much stable situations.

Defence

A single decision can turn happy walking people to dead carrying bodies. In defence sector or in intelligent service make right decision plays a huge role. Winning the war would be very easy if there were no casualties or environmental challenges. But as a country has to be ready for these situation and to take right decision in time we need a lot of information which we have such as vehicles we use in defence and the equipments information's of the expected countries human power details their strength and weakness same with other countries strength and weakness the support other country having and there power strength and weakness along with their limitation in participation. Other major factors such as historical actions taken by the county at different situations. Basically all this will turn into a lot of data that's our Bigdata. To have right decision we need to do the analysis on this the insight we get could help defence

authorities in taking decision which can cause minimum casualties but more effective strategy for winning the war.

Natural Disasters

The world faced significant number of natural disasters like earthquakes, floods, tsunami, cyclones, volcanoes etc. These cost a planet a lot like lives, health and ofcourse economics. We could avoid such a huge loss or at least minimize it if we could predict it earlier. This could help in saving thousands of animals and human lives. For these to happen we must have disaster system design which includes hardware and software. In this system it shall take lot of sensors and it produces the lot of data. Sensor always produce the un structural data which can be received using big data tools and they also produce huge amount of data that also can be consumed by the big data tools. With these data we can be able to get the insight and predict the unwanted disasters to come in early stage which in turn help organizations to take early steps in avoiding casualties and minimizing the economical cost.

III. PRIVACY AND SECURITY CONCERNS

The more information is collected up by powerful computers as giant sets of data, big data, the more likely that it includes personal or sensitive information. It is important to remember that cyber security is to be as strong as the weakest link in its chain. If one of these persons lacked adequate security controls and suffered a data breach, the data they're trying to protect between them would now be vulnerable. Cybercriminals are now focusing privileged users to access economic data. Therefore the strict enforcement of admission policies and continuous monitoring of activities to detect anomalous activities are very important. Strong encryption is vital to prevent data burglary, alteration, discovery or destruction in the entire life cycle of data, especially during storage and transmission. Employees should be aware of various attacks like spear phishing attack that target to open malicious attachments or click infected links. Kaspersky Lab published a massive cyber-attack targeting the financial/banking industry [11]. Now the insiders are responsible for more financial sector attacks than outsiders. Security events to be identified by correlation and analytical tools as a malicious attack that will gather, interrupt, repudiate, damage or terminate information system resources or information itself. Protecting Big Data privacy is also an important area to consider. Individuals and institutions who entrust their data to an organization expect their private information to be protected. However, recent surveys of the American public suggest that there is low confidence in the ability of organizations to guarantee privacy of their data. Notably, only 38% of the respondents stated that have confidence that their credit card companies will ensure privacy and security of their personal credit card activity records. This proportion of confident respondents drops to 16% and 11%, respectively, when asked about their trust of search engine providers and social media sites [7]. As personal information becomes increasingly shared, bought and sold as a commodity, incidents of unauthorized disclosure are likely to grow. Such incidents will further erode public

confidence in privacy of their information. While a major concern, identity theft by cyber intrusion is only one of the privacy challenges faced by the public in the era of Big Data. This era raises new privacy risks that have been collectively denoted the “4 R challenges of Big Data” in [7]: Reuse, Repurposing, Recombination, and Reanalysis. Reuse, repurposing and reanalysis risks arise when data that is collected with prior consent for one purpose are analyzed for another purpose that may cause harm to the individual. Recombination denotes the risk of personal reidentification achieved by combining collected data with information from other data sources. By launching an “inference attack” on the data, recombination can succeed, at least partially, even when the data is summarized, anonymized and encrypted.” Recent research on new methods of privacy protection, e.g., differential, shows that an individual’s information can be protected from recombination with some loss in data fidelity. However, associated losses in data quality may decrease data utility for the organization and its data-sharing partners. This raises several questions. 1) What are the “privacy-at-risk” data sharing practices in the investment, commercial banking and financial services sectors? 2) How can these risks be reduced and consumer confidence to be improved? 3) Should the banking industry consider adopting IRB-style broad consent regulations governing the use and sharing of personal data? 4) Should the consumer have more control on how their financial data will be anonymized, used and shared? 5) What new categories of personal data are emerging, e.g., personal accessory data (IoT) and health data (wearable sensors), that could be properly used or shared? [7]

IV. CONCLUSION

All the applications presented by this paper need right and quality of input to get the right outcome. As the data is growing it has become challenging day by day. Even growth of data seem to be advantage the verities of the data lead to poor input feeding right data will only give the best outcome. It has become challenging in data cleaning process which has a larger effect on any algorithm used. In case of successful implementations today’s tools not just give predictive analysis it also give in real-time. This is being used in every area we can name. We have discussed different application use case some saves lives then some saves business and some protects the law and enforce by fraud detecting. There is need of improvement in critical fields such as privacy and security of sensitive information which can be enhanced with different techniques available such as encryption of the data at different level. And strict access control to these systems and implementing additional layer of securities which are best in industries.

REFERENCES

[1] Feng Xia, Senior Member, IEEE, Wei Wang, Teshome Mergers Bekele, and Huan Liu, Fellow, IEEE, “Big Scholarly Data: A Survey”, IEEE Transaction on Big Data, Vol. 3, No. 1, January-March 2017.

- [2] Deepankar Bharadwaj, Research Scholar, Dr. ArvindShukla, IFTM University. HOD, Department of Computer Applications, IFTM University, Moradabad(UP), “Text Mining Technique on Big Data Using Genetic Algorithm (A Review)” , International Journal of Computer Engineering and Applications, Volume X, Issue IX, Sep. 16, Issn 2321-3469.
- [3] Wei Tan, M. Brian Blake and I. Saleh, Schahram Dustdar, “Social-Network Sourced Big Data Analytics”, IEEE INTERNET COMPUTING, 10897801/13/\$31.00 © 2013 IEEE
- [4] Matthew Smith, Benjamin Henne, , “Big Data Privacy Issues in Public Social Media”, 978-1-4673-1703-0/12/\$31.00 ©2013 IEEE
- [5] Monreale et al., “Privacy-by-design in big data analytics and social mining”, Data Science 2014, 2014:10
- [6] Shen Yin, Okay Kaynak, “Big Data for present-day Industry: Challenges and Trends”, Proceedings of the IEEE, Vol. 103, No. 2, February 2015.
- [7] Alfred O. Hero and R. Jamison and Betty Williams Professor of Engineering, University of Michigan; Co Director, Michigan Institute for Data Science (MIDAS) , Data Privacy and Security, Big Data on Finance, Center on Finance, Law and Policy, University of Michigan Law School, October 27, 2016
- [8] Akinul Islam Jony, Department of Computer Science, American International University – Bangladesh, Applications of Real-Time Big Data Analytics, International Journal of Computer Applications (0975 - 8887), Volume 144 - No.5, June 2016.
- [9] Khantil Choksi,Niriksha Dalal, Mr.Kshitij Gupte and Dr.Anjali Jivani, “Security And Privacy Challenges In Big Data”, International Journal of Latest Trends in Engineering and Technology Vol.(7) Issue(3), pp. 313-318.
- [10] Doug Olenick, Online Editor, SC Media US, Cybercrime, Financial services sector most attacked in 2016: IBM, [https:// financial-services-sector-most-attacked-in-2016-ibm/article/ 653706/](https://financial-services-sector-most-attacked-in-2016-ibm/article/653706/), April 28, 2017
- [11] Fran Howarth, Sabotage: The Latest Threat to the Financial/Banking Industry, “<https://securityintelligence.com/sabotagethe-latest-threat-to-the-financialbankingindustry/>”, August 30, 2016

- [12] Krishna Kumar Singh, Research Scholar, Priti Dimri, PhD, Associate Professor and Head, Krishna Nand Rastogi Research Scholar, Dept. of CS G.B.P.E.C, Pauri Garhwal Uttarakhand, “The Implications of Big Data in Indian Stock Market” in 2014.
- [13] Neil Couch and Bill Robins, “Big Data for Defence and Security”, Royal United Services Institute, Occasional Paper, September 2013
- [14] Lech Janczewski, Khairulliza Ahmad Salleh ‘Technological, organizational and environmental security and the privacy issues of the big data:- The publicity review’, Science Direct, Procedia Computer Science 100 (2016) 19 to 80.
- [15] Jason Parms, [https:// www. Business. com / articles /privacy – and – security -issues – in – the – age – of – big - data](https://www.Business.com/articles/privacy-and-security-issues-in-the-age-of-big-data), February 22, 2017
- [16] Jung woo Ryoo, Big data security problems threaten consumers’ privacy, March 23, 2016 9.08pm AEDT