# SAFEGUARDING PRIVACY IN BIG DAta

Dr. D. USHA, M.C.A., M.Phil., M.Tech., Ph. D

**Abstract**— Enhanced connectivity and computing power, as well as increased uptake of big data technologies, not only provide opportunities for organizations but also give rise to new challenges in terms of privacy and security. Their degree introduced by analytics vary depending on the kind of processed data. Though gathering and processing of vast amounts of data is not new, the speed of processing is new. Big data is currently a major topic of discussion across many areas like scientific research, national security, government data and management. Public and private sectors use big data analytics as even long running queries are now be resolved in a tenth of a second with terabytes of directly addressable RAM. Data streaming from various sources is mostly impersonal, but the digital economy's increasing reliance on trade of personal information is an issue. There is fear that these growing markets may infringe individual's rights to privacy and not protect their data. A record seven administrative proceedings and actions were taken by the Federal Trade Commission in 2014 [1], as contending companies failed to provide appropriate security for consumers' personal information. These stimulated discussions between policy makers, regulators and specialists across these policy areas. The challenge is in identifying data collection that violates privacy. This paper proposes a few guidelines on privacy protection big data.

**Index Terms**—Big Data, Big Data Analytics, Application Encription, Cyber Attack, Privacy Protection Management, big Data Privacy.

## 1 INTRODUCTION

Current people, devices and networks constantly generate data, even if the devices are not used the networks are busy generating its location and relevant data. The mobile data traffic is also growing rapidly. Smartphone subscriptions is expected to be over 6 billion in the coming years with five times the traffic of today [2]. Big data consists of many types of data from multiple sources. The big-data-driven telecom analytics market alone is expected to grow by fifty percent and reach revenues up to USD 5.4 billion at the end of 2019 [3]. This superficial impact will also create business and opportunities for organizations in new areas. Improved real-time connectivity and data management will create tailored data for analysis and learning, thus enabling improvements in many business areas like transport, logistics, energy, environmental monitoring and agriculture. Organizations, prevented from benefiting from the value of big data in the past due to its volume, velocity and variety, will be facilitated by accurate and updated data. The three V's depicted in Fig 1.

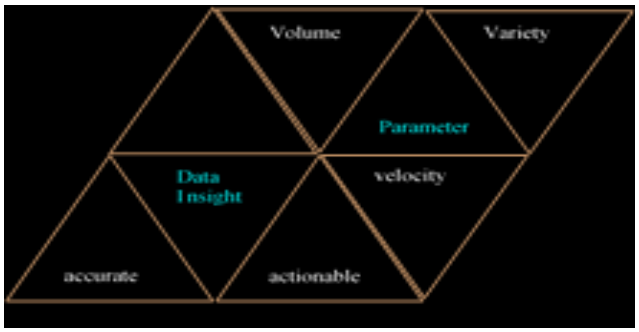


Fig. 1: The V's and A's of in big data

For organizations, big data entails deploying powerful real-time analysis of data analytics, cooperative processing and the ability to automatically use existing applications vital to the organization's survival. In the present scenario, every organization relies on processing collected data about customers or employees. Organizations can make use of big data to drive a wide range of important decisions like creating and recommending competitive offers to customers, communicating with users about their usability, delivering more reliable services and monitoring to proactively solve problems. Thus creating outcomes like improved user experience, increased customer satisfaction and a healthy progress into the Networked Society. Protecting collected information in big data is vital. Storing and processing this data has made data protection a major risk-management issue. Organizations need to be transparent in data collection and explain compliance with data protection principles. A data breach or an attack can cost heavy losses in revenue, software, litigation fees, forensic analysis costs, to name a few. Thus Data protection needs to be a priority for the entire organization and be an overall part of corporate culture. Policies should be in place on collection, period of storage and destruction of data. Studies have found that many companies failed to practicing basic security and above ninety percent of the breaches were preventable with encryption, secure data backup and data access control. One key to personal data protection is to ensure a fair processing, especially in decisions affecting individuals. The complexity of big data analytics cannot be an excuse obtaining required consent from users. By scrutinizing data in real time, systems can identify potential sources of threats. Cyber insurance offers a valuable safeguard from the financial damage that a data breach can impose on a company, but the losses may not be insurable. Compliance is a major challenge for both the companies and the insurance industry. Creating best practices for security on all points of access to network operations can prevent cyber interferences. By putting rules in place and creating a culture of cyber preparedness and responsiveness, they will be much more capable of controlling risk and keeping data in IT systems safe and secure [4]. High-profile security gaps with hacking using with smart devices, raise consumer concerns about their personal information's privacy and security. Nevertheless, big data and analytics are key tools for success with connected home products and systems. This paper is intended to give an overview of the issues observed and contribute to the debate on big data and privacy. The paper aims is to ensure that the different privacy risks of big data are con-

sidered along with the benefits of big data as the benefits cannot simply be traded with privacy rights.

• *Dr. D. Usha is currently working as a Assistant Professor in Mother Teresa Womens's University, Kodaikanal, Tamilnadu India phone (91) 04542-241685, +91 8508779282*
• *E-mail: ushadanabal@gmail.com*

## 2 Security issues in big Data

Big data presents a remarkable opportunity for scientists, product managers, marketers and enterprises in tapping into new volumes and varieties of data. Without the right security and encryption solution in place, however, big data can mean big problems. The security challenges in big data is detailed below.

### 2.1 Source and Frameworks

The advantages of big data is exploited by organizations in various forms including a range of heterogeneous applications like resource planning systems, video files, social media feeds and spreadsheets to name a few. New data sources are also added and unknown variants of data may be added in future. These sources can include personally identifiable and important information, intellectual property and much more. This data needs to be secured by addressing security policies and compliances. Within the big data environment sensitive data may be processed anytime.

### 2.1 Analytics

Output presented in dashboards and reports, is the ultimate fruit of a big data initiative. It stems from analytics that help business innovations and optimizations. Visa is looking at using big data analytics to develop authorization of credit card payments in a new way [5]. The analytics may present a sensitive asset as a critical differentiator, but can fall into the wrong hands. Big data is also valuable to cyber criminals or a disgruntled system administrator looking to make quick and illicit money for its attributes. Establishing effective security across is both critical and challenging in big data. Organizations are leveraging on cloud-based services to support their big data resources where the task of managing security becomes even more difficult. Users have to contend with the vendor's infrastructure security with a potential exposure to others and a number of other additional risks. Big data analytics can involve repurposing personal data. Fig. 2 depicts the Big data analytics stack

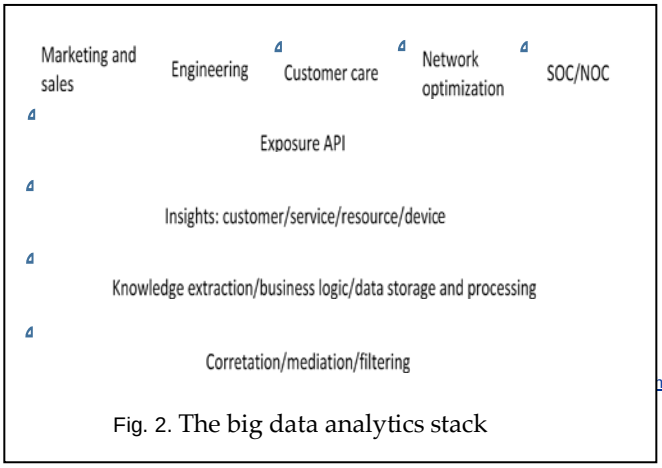


Fig. 2. The big data analytics stack

## 3 Securing Big Data

Compliant security solutions in big data can maximize the benefits of big data analytic. The Security Data Security Platform can offer microscopic controls, strong encryption, and complete coverage. The organizations can secure sensitive data including big data infrastructure, and big data analytic results. Security enabled security teams can leverage centralized controls that optimize efficiency and compliance adherence, while the Security Platforms can offer encryption and access control. Organizations can leverage on both structured and unstructured data, for their big data initiatives [6]. Data from databases, data warehouses, system logs, spreadsheets, and many other diverse systems may become a part of big data, but to establish security for these diverse data sources, organizations can use Security Transparent Encryption and control access at the file-system level. The encryption solution should be easy to deploy without any changes to existing applications. A Security Application Encryption can encrypt specific columns in an application and by encrypting a specific column or a sensitive field, the information will remain incomprehensible within the big data environment.

### 3.1 Securing Frameworks and Analytics

In big data environments, replication and migration of data occur amongst a large number of nodes along with sensitive information present in system logs, configuration files, disk caches and error logs. A Transparent security Encryption can efficiently protect data across these areas, delivering encryption, privileged user access Use of algorithms. Before the advent of big data, analyzing a dataset involved, constructing a query to find it, by identifying the relevant entries. Big data analytics, on the other hand, often involves running a very large number of algorithms against the data in order to find correlations [7], rather than testing a particular hypothesis. Once relevant correlations have been identified, a new algorithm can be created and applied to particular cases. This is a form of 'machine learning', since the system 'learns' which are the relevant criteria from analyzing the data. While algorithms themselves are by no means a new concept, their use in this way is a feature of big data analytics, control and security intelligence. Big data output, an intellectual property of an organization which comes in many output forms, run the risk of being attacked. Security Transparent Encryption can be deployed on servers to provide security for these confidential big data assets. It can encrypt big data outputs and control and monitor accesses. An Application Encryption can be used to encrypt data from a specific application or field. Some practical aspects of big data analytics are summarized in Table 1

TABLE 1
ASPECTS IN BIG DATA ANALYT-

| | |
|---|---|
| Personal Data | All of big data project need not use personal data. Personal data can be anonymised? Data usage can be evaluated for security compliances |
| Impact | Organizations can carry out a privacy impact assessment |
| Repurposing Data | New purposes can be checked for compatibility in privacy and protection terms. While buying personal data, due diligence has to be practiced and protection of data has to be ensured while processing personal data. |
| Transparency | Organizations have to be transparent and open on operations while explaining the purposes in implications and benefits of the analytics. Innovative and effective ways of assurances on privacy protection can be conveyed to the people concerned. |
| Data minimization | Big data analytics is not an excuse for piling stocks of data or keeping it longer than required period. Long term uses must be articulated and justified when future use is undefinable, while guaranteeing security of private information. |

ICS

### 3.2 Safeguarding Personal Information

sing neighbor discovery protocol, each sensor node knows its Privacy preservations require new techniques, as security challenges can be overcome. New challenges include protection against personal data compiled from different sources, as well as the necessary changes to privacy-enhancing technologies based on the nature of data. Existing legislation may not be enough to protect user privacy and may soon become outdated in evolving technology environments. Regulations also may vary in their approach based on big data technology and country. These privacy challenges can be addressed by increasing awareness in working and taking privacy considerations in design of a new product or service [8]. This includes data life cycle management at each stage. Internal processes can be adjusted by performing privacy audits. Standardization activities such as the introduction of the ISO 29100 standard [9] can also support internal work. Technical approaches to address security and privacy include established security techniques like encryption in secure communication or data storage and severe access including techniques like collection, quality and retention in data, logging data, audits and anonymization methods [10]. Figure 3 depicts Creating value from Big Data.

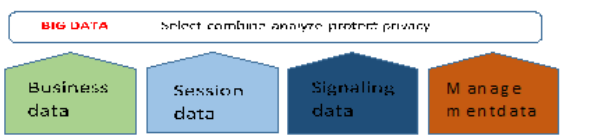


Fig. 3. Creating value from Big Data

Companies are safeguarding personal information mainly due to the reputational damage and liability from data breaches. They need to comply with standards and rules in safeguarding privacy. Companies that have been successful in escaping government actions are those that evolved in planning and executing information security on sensitive personal information with a perfect response after a cyber-attack. Rather than focusing on big data technology, organizations should start from the business value they want to create and apply extensive competence to understand insights into raw data before applying big data techniques. Any decision driven by analytics which is accurate and timely has a better chance to profit from change. The key to effective decisions is the ability to combine data from several sources and examine large amounts of data for a comprehensive view of the business. Even unimportant data can reveal new insights like combined data from social media can help understand users experience of a service or in knowing whether they like or dislike. Organizations need to understand key data relationships like complex hierarchies and links between data types and sources. The financial services industry has been rather stringent in securing client data. Humanizing and Safeguarding big data can be done by a two-factor authentication with two separate passwords to access or use can provide greater protection. Business owners can learn best practices to help protect their data as repairing the damage from a data breach can ruin a business. Managing big data requires planning, discipline and vigilance. This section lists a few strategies to organize and protect big data. Businesses can use the network of computer servers in a cloud infrastructure, since cloud services are secured. Communicating with a mobile number as an access point can help one's identity, acting as a virtual fingerprint for identification. Sensors in the street or in shops can capture the unique MAC address of the mobile phones of people passing by [11]. Although the MAC address does not itself identify a specific individual, it could be used to track repeated visits. The most secretive like personal mobiles and ATM card numbers for access can protect valuable resources. A communication can be interrupted for user inputs for privacy. Though entering a code after a few minutes is annoying, it can prevent others from gathering valuable data or committing fraud on personal information. When phone data is lost or changed, a third party interaction like an email can help restore the privacy settings, Encrypting personal data with software tools at reasonable costs can help, as keys are required to retrieve information. Customers or users should send encrypted personal information. Data protection is concerned with personal data, but it is important to remember that many instances of big data analytics do not involve personal data at all. Examples of non-personal big data include: world climate and weather data; using geospatial data from GPS-equipped buses to predict arrival times; data from radio telescopes in the Square. Moreover, there are many examples of big data processing of personal data like medical patients data or data on purchases. If personal data becomes impersonal when anonymised and ceases to be personal data and becomes difficult to identify individuals in data or data in combination with other data used to identify them.

## 4 CONCLUSION

Storing and processing volumes of data is no longer an issue. This could enable the creation of a range of user-centric applications and services with benefits. An analytics platform

which is big and horizontal is necessary to support a variety of applications like analysis of real incoming data, data correlation with domain expertise and exposing insights into data. True data-driven insight calls for domain expertise. For organizations this means in-depth knowledge of network functions, what data to pull from network's nodes and connecting data from multiple sources to yield an enriched information. This approach can both enhance the performance of applications justifying big data investments. The big data tools and technologies need to support finding insights that are actionable, accurate and adequate. Big data, organizations should focus their attention towards privacy protection in the three A's of big data. The conclusion is that though people are sharing their personal data, this does not necessarily mean that data protection is obsolete. Data protection does indeed have an important role to play in the online environment, probably even more so than in the physical world context. Thus data protection is not only still relevant but will be crucial to address the potential harms of big data and analytics and there is no doubt that privacy protection has a role to play in big data.

## REFERENCES

[1] 1] http://www.dandodiary.com/2014/04/articles/cyber-liability/district-court-upholds-ftcs-authority-to-bring-data-breach-enforcement-action

[2] [2] Ericsson, Ericsson Mobility Report, February 2015, available at: http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-feb-2015-interim.pdf

[3] [3] Research and Markets, Carrier B2B Data Revenue: Big Data, Analytics, Telecom APIs, and Data as a Service (DaaS) 2015-2020, July 2015, available at: http://www.researchandmarkets.com/reports/3071341/carrier-b2b-data-revenue-big-data-analytics.

[4] [4] Cmdr. Dave Pettinari, Pueblo County Sheriff's Office, Framework for Conducting an Investigation of a Computer Security Incident, Standard Operating Procedure -- Pueblo High-Tech Crimes Unit Investigative Protocol -- Hacking and Intrusions, Apr 2000 Investigating Cyber Crime/Hacking and Intrusions

[5] [5] The future of technology and payments. Edition 2. Visa Europe, April 2013 http://www.visaeurope.com/en/about_us/industry_insights/tech_trends.aspx Accessed 25 June 2014

[6] [6] Russom, Philip Managing big data. The Data Warehousing Institute, 2013. Available from http://www.pentaho.com/resources Accessed 25 June 2014

[7] [7] Centre for Information Policy Leadership. Big data and analytics. Seeking foundations for effective privacy guidance. Hunton and Williams LLP, February 2013 http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Anal ytics_February_2013.pdf Accessed 25 June 2014

[8] [8] A. Cavoukian, Privacy by Design – The 7 Foundational Principles, available at: https://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles

[9] [9] ISO/IEC, 29100:2011 Information technology – Security techniques – Privacy framework, 2011, available at: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=45123

[10] [10] http://www.ericsson.com/ph/res/docs/whitepapers/wp-big-data.pdf

[11] [11] Seward, Zachary M and Datoo, Siraj City of London halts recycling bins tracking phones of passers-by Quartz 12 August 2013 http://qz.com/114174/city-of- london-halts-recycling-bins-tracking-phones-of-passers-by/ Accessed 25 June 2014.