# BIG DATA- SECURITY AND PRIVACY

**Ritika Chugh** [1]

[1]Assistant Professor,CSE,

Vaish College of Engineering, Rohtak (Haryana),

ritikachugh92@gmail.com


**Pooja** [2]

[2]Assistant Professor, Department of Computer Application (BCA),

Vaish College of Engineering, Rohtak (Haryana)

balhara.pooja2@gmail.com

**Abstract**

The voluminous amount of data i.e Big Data is rising day by day in the world by using internet ,social networking sites, cell phones and many other advance technology. In big data research ,security and privacy plays vital role which can be characterized by 3V's: Volume, Variety and Velocity . Big data is extremely large data sets which may be analyzed computationally for revealing the trends , associations and patterns which relates to human behaviour and interactions . In this paper, privacy is reserved using some existing techniques such as privacy preserving aggregation , de-identification and operations over encrypted data.

Keywords : big data, Privacy preserving techniques, Velocity, Volume, Variety, IaaS, PaaS , SaaS

**Introduction**

The term 'Big Data' is used for voluminous amount of data sets which have complex structure. Big data is categorized into Structured and Unstructured data. Privacy and security are major problems which are concerned in big data. Privacy refers to ability to decide what individual information goes where. This will help in usage of user's information in appropriate manner. Privacy is safety of information from other people . While the transmission of personal

information over the internet, privacy is applied. Security refers to protection of data from mischievous attacks , avoid the misuse of stolen data and also keeps the data confidential. Big data plays vital role in cloud computing. Network accessible resources are defined in cloud computing as services and these services are delivered using service model of cloud computing .

Three V's of Big data are

1. Volume: It determines the quantity of data and tells about the size and potential of data. It also refers to exponential growth in the volume of data. Various form of data are music, videos, large image files, text data etc. Nowadays data can be stored in terms of terabytes and petabytes depending upon different enterprises.
2. Variety: It refers to different forms of data i.e Structured and unstructured data. Structured data includes numbers, dates and group of words etc. Unstructured data includes e-mail messages, documents, videos, photos, audio files etc.
3. Velocity: It refers to the speed at which data is generated or processed to meet the requirements. Biggest challenge for many organizations is to deal with data velocity because of its high speed.
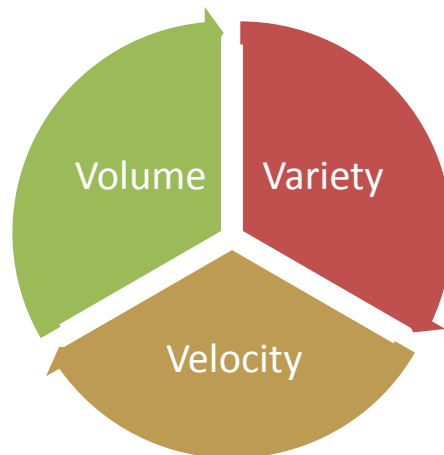


Fig.1: 3V's of Big Data

Types of Big data

Big data is characterized into structured and unstructured data.

Structured data: Structured data are words and numbers which can be easily categorized and analyzed. It also includes account balance, transaction data etc and are generated by network sensors which are embedded in smart phones, GPS(Global Positioning System) devices and electronic devices.

Unstructured data: Unstructured data can't be categorized and analyzed. It includes e-mail messages, documents, videos, photos, audio files etc and also include more complex information.

Existing Privacy Preserving Techniques:

1. Privacy preserving aggregation: On some homomorphic encryption, privacy preserving aggregation is built which is a popular data collecting technique for event statistics. In this phase of big data collecting and storing it can protect individual privacy. It is purpose specific. One purpose aggregated data can't used for other purposes. Given a homorphic public key encryption algorithm E(.), different sources can use the same public key to encrypt their individual data $m_1$, $m_2$.....,$m_n$ into ciphertext $c_1$=E($m_1$), $c_2$=E($m_2$)...... $c_n$=E($m_n$). By taking sum aggregation as an example these ciphertext can be aggregated as C=$\pi_i^n$ = 1$c_i$=E($\sum_i^n$=1$m_i$). The aggregated result $\sum_i^n$=1$m_i$ can be recovered from C.

2. Operations over encrypted data: It has been widely studied in cloud computing for keeping the sensitive documents private and their keywords which are encrypted and stored in a cloud server. Set of encrypted documents are returned by the server when user submits 'capability' encoding query conditions. It is complex and time consuming.

3. De-identification: For privacy preserving , data mining and traditional technique is used which is de-identification but it is not sufficient for protecting big data privacy.

Before the release for data mining data should be first sanitized with generalization and suppression. If we compare privacy preserving aggregation and operations over encrypted data then de-identification can make data analytic and mining more flexible and effective.

Big data security techniques

1. Application Software Security
2. Monitoring , Maintenance and Analysis of Audit Logs
3. Software Configuration for Software and Hardware
4. Account Monitoring and control

**Conclusion**

This paper presents the overview of 3V's of Big data i.e Volume , Variety and Velocity. Big data is characterized into structured and unstructured data. This paper includes Existing Privacy Preserving techniques and if we compare all these techniques , we concluded that de-identification can make data analytic and mining more effective and flexible. This paper also presents the Big data security techniques.
Hence privacy and security , both are very important for Big data

**References**

1  Asha Patel "A Survey Paper on Security Issue with Big Data on Association Rule Mining", IJIRST, National Conference on Latest Trends in Networking and Cyber Security, March 2017

2  Minit Arora, Dr Himanshu Bahuguna , "Big Data Security – The Big Challenge", International Journal of Scientific & Engineering Research, Volume 7, Issue 12, December-2016,ISSN 2229-5518

3  Ninny Bhogal , Shaveta Jain ," A REVIEW ON BIG DATA SECURITY AND HANDLING", International Research Based Journal ,Vol(6)-Issue(1), March,11, 2017,ISSN 2348-1943

4   Tilwani Mashook ,Patel Malay, Pooja Mehta, "Security and Privacy –A Big Concern in Big Data A Case Study on Tracking and Monitoring System", IJIRST, National Conference on Latest Trends in Networking and Cyber Security, March 2017