# Chapter 13
# Security and Privacy in Big Data Computing: Concepts, Techniques, and Research Challenges

**Kiritkumar J. Modi**
https://orcid.org/0000-0001-6462-059X
*Parul University, India*

**Prachi Devangbhai Shah**
*U. V. Patel College of Engineering, Ganpat University, India*

**Zalak Prajapati**
*U. V. Patel College of Engineering, Ganpat University, India*

## ABSTRACT

*The rapid growth of digitization in the present era leads to an exponential increase of information which demands the need of a Big Data paradigm. Big Data denotes complex, unstructured, massive, heterogeneous type data. The Big Data is essential to the success in many applications; however, it has a major setback regarding security and privacy issues. These issues arise because the Big Data is scattered over a distributed system by various users. The security of Big Data relates to all the solutions and measures to prevent the data from threats and malicious activities. Privacy prevails when it comes to processing personal data, while security means protecting information assets from unauthorized access. The existence of cloud computing and cloud data storage have been predecessor and conciliator of emergence of Big Data computing. This article highlights open issues related to traditional techniques of Big Data privacy and security. Moreover, it also illustrates a comprehensive overview of possible security techniques and future directions addressing Big Data privacy and security issues.*

DOI: 10.4018/978-1-7998-8954-0.ch013

## INTRODUCTION

In the era of distributed computing data are scattered among different machine. The rapid and exponential growth of data has increased the storage size where we can store huge pile amount of data. As per the Google report, 2.5 quintillion units of data are generated per day and this data is coming from different sources like social media, banking sector, Internet of Things, mobile generated data, etc. Data is very crucial part in any sector for communication and Information. This all data in form of structured, unstructured and semi structured type so we need to provide security on this data to achieve confidentiality. There are four basic attributes that defines Big Data, which are known as four V's: volume, variety, velocity, and veracity. The main trait that makes data "big" is its sheer volume. Due to digitization, continuous feeding of unstructured data flows from various sources and thus variety of data increases. In this era structured data is easily augmented by unstructured data. Veracity refers to the reliability of the data. Accuracy and trustworthiness of data is measured through veracity factor. Velocity is the rate at which the huge amount of data that is generated and needs to be processed.

The security of big data relates to all the solutions and measures to prevent the data from threats and malicious activities. Security refers to personal freedom from external forces. The main objective of security are confidentiality, integrity, and availability. Moreover, privacy is one's right to freedom from intrusion. Privacy prevails when it comes to processing personal data, while security means protecting information assets from unauthorized access (Mahmood & Afzal, 2013).

Higher Integrity and confidentiality can be achieved by providing security on three levels. First level is data storage level where crucial and important information stored e.g., credit card information, customer information. The Second level is built as a strong big data security tool e.g. a firewall, which can prevent unauthorized user to access information by filtering traffic. Third level is Implementing Access control method, which can access data by centralized key management. By developing policies, procedures and security software, it is possible to protect data at every level by against malware and unauthorized access (Gahi, Guennoun & Mouftah, 2016).

Cloud computing is the commodification of computing and data storage by means of globally accepted techniques. The advantages of having big data on cloud are cost cutting, availability of instant infrastructure and faster access of data. The integration of big data with cloud storage also leads to many privacy breaches. One of the reasons for these breaches is that no appropriate security application is available to achieve privacy goals for such massive data. The shifting towards big data in the cloud has many benefits; it can bring powerful data analytics and boost decision making in data driven approaches. Cloud-based data analytics requires high-level, easy-to-use design tools for dealing with huge, distributed data sources.

The 2018 Thales Data Threat Report (DTR) (Mahmood & Afzal, 2013) surveyed 99% organization uses big data with security techniques e.g. Stronger authentication and access controls, Improved monitoring and reporting tools (Jain, Gyanchandani, & Khare, 2016), Encryption and access controls for underlying platforms (Jain, Gyanchandani, & Khare, 2016).

In this chapter we are going to discuss different Encryption technique and Key management technique and compare it. These two techniques are used to provide security at Storage level and Access Control level to protect system from Ransomware, Distributed Denial of service attack threats which can crash server or leak sensitive information. We elaborate different security techniques with their pros and cons.