Chapter 3
Big Data Security: Challenges, Recommendations and Solutions

Fatima-Zahra Benjelloun
Ibn Tofail University, Morocco

Ayoub Ait Lahcen
Ibn Tofail University, Morocco

ABSTRACT

The value of Big Data is now being recognized by many industries and governments. The efficient mining of Big Data enables to improve the competitive advantage of companies and to add value for many social and economic sectors. In fact, important projects with huge investments were launched by several governments to extract the maximum benefit from Big Data. The private sector has also deployed important efforts to maximize profits and optimize resources. However, Big Data sharing brings new information security and privacy issues. Traditional technologies and methods are no longer appropriate and lack of performance when applied in Big Data context. This chapter presents Big Data security challenges and a state of the art in methods, mechanisms and solutions used to protect data-intensive information systems.

INTRODUCTION

The value of Big Data is now being recognized by many industries and governments. In fact, the efficient mining of Big Data enables to improve the competitive advantage and to add value for many sectors (economic, social, medical, scientific research and so on).

Big Data is mainly defined by its *3Vs* fundamental characteristics. The *3Vs* include *Velocity* (data are growing and changing in a rapid way), *Variety* (data come in different and multiple formats) and *Volume* (huge amount of data is generated every second) (Wu, Zhu, Wu, & Ding, 2014). According to (Berman, 2013) these three characteristics must coexist to confirm that a source is a Big Data source. If one of these three *Vs* does not apply, we cannot discuss about Big Data.

(Berman, 2013) and (Katal, Wazid, & Goudar, 2013) indicate that more *Vs* and other characteristics have been added by some Big Data actors to better define it: *Vision* (the defined purpose of Big Data

DOI: 10.4018/978-1-5225-7501-6.ch003

mining), *Verification* (processed data comply to some specifications), *Validation* (the purpose is fulfilled), *Value* (pertinent information can be extracted for the benefit of many sectors), *Complexity* (it is difficult to organize and analyse Big data because of evolving data relationships) and *Immutability* (collected and stored Big Data can be permanent if well managed).

Beside this, some argue when defining Big Data, that any huge amount of digital data sets that we can no longer collect and process adequately, through the existing infrastructures and technologies, are by nature Big Data.

In this chapter, we are interested in security challenges faced in Big Data context. We present also a state of the art in several methods, mechanisms and solutions used to protect information systems that handle large data sets.

Big Data security has many common points with the security of traditional information systems (where data are structured). However, Big Data security requires more powerful tools, appropriate methods and advanced technologies for rapid data analysis. It requires also a new security management model that handles in parallel internal data (data produced by internal systems and processes within an organization) and external data (e.g., data collected from other companies or external web sites). Regarding those points, many questions can be raised: i) How to manage and process securely large, unstructured and heterogeneous types of data sets? ii) How to integrate security mechanisms into distributed platforms while ensuring a good performance level (e.g., efficient storage, rapid processing and real-time analysis)? iii) How to analyse massive data streams without compromising data confidentiality and privacy?

This chapter presents first these challenges in detail. Then, it discusses various solutions and recommendations proposed to protect data-intensive information systems.

SECURITY CHALLENGES IN BIG DATA CONTEXT

As mentioned by (Kim, Kim, & Chung, 2013), security in Big Data context includes three main aspects: information security, security monitoring and data security. For (Lu et al., 2013), managing security in a distributed environment means to ensure Big Data management, system integrity and cyberspace security.

Generally, Big Data security aims to ensure a real-time monitoring to detect vulnerabilities, security threats and abnormal behaviours; a granular role-based access control; a robust protection of confidential information and a generation of security performance indicators. It supports rapid decision-making in a security incident case. The following sections identify and explain a number of challenges to achieve these goals.

Big Data Nature

Because of Big Data velocity and huge volumes, it is difficult to protect all data. Indeed, adding security layers may slow system performances and affect dynamic analysis. Thus, access control and data protection are two “BIG” security problems (Kim et al., 2013). Furthermore, it is difficult to handle data classification and management of large digital disparate sources. Even though that the cost by GB has diminished, Big Data security requires important investments. In addition to all that, Big Data is most of the time stored and transferred across multiple Clouds and distributed worldwide systems. Sharing data over many networks increase security risks.