Chapter 2
Big Data:
Challenges and Solutions

P. Lalitha Surya Kumari
Geethanjali College of Engineering and Technology, India

ABSTRACT

This chapter gives information about the most important aspects in how computing infrastructures should be configured and intelligently managed to fulfill the most notably security aspects required by big data applications. Big data is one area where we can store, extract, and process a large amount of data. All these data are very often unstructured. Using big data, security functions are required to work over the heterogeneous composition of diverse hardware, operating systems, and network domains. A clearly defined security boundary like firewalls and demilitarized zones (DMZs), conventional security solutions, are not effective for big data as it expands with the help of public clouds. This chapter discusses the different concepts like characteristics, risks, life cycle, and data collection of big data, map reduce components, issues and challenges in big data, cloud secure alliance, approaches to solve security issues, introduction of cybercrime, YARN, and Hadoop components.

INTRODUCTION

This chapter updates the most important characteristic about how computing framework must be configured and intelligently administers to fulfill the main aspects related to security required by the applications of the Big Data. The place where the huge quantity of data is stored, extracted and processed is called Big Data. Big Data is the area with vast potential to handle datasets of the size which is beyond the capability of generally used software tools to capture, manage, and timely analyze that amount of data. The amount of data to be analyzed is estimated to be twice for every two years. The various sources of this unstructured data are medical records, scientific applications, sensors, social media, video and image archives, surveillance, business transactions, Internet search indexing, and system logs. As the number of connecting devices of Big Data like Internet of things increases attention towards Big Data also increased at unexpected levels. In addition, it is very common to acquire on-demand supplementary computing control and storage to carry out exhaustive data-parallel processing from public cloud providers. Thus,

DOI: 10.4018/978-1-5225-9742-1.ch002

privacy and security issues can be potentially increased by the variety, volume and wide area deployment of the system infrastructure to support Big Data applications. As Big Data expands with the help of public clouds, traditional security solutions tailored to private computing infrastructures, confined to a well-defined security perimeter, such as firewalls and demilitarized zones (DMZs) are no more effective. Using Big Data, security functions are required to work over the heterogeneous composition of diverse hardware, operating systems, and network domains. A clearly defined security boundary like firewalls and demilitarized zones (DMZs), conventional security solutions adapted to private computing infrastructures are no more effective for Big Data as it expands with the help of public clouds.

Using Big Data, security functions are required to work over the heterogeneous composition of diverse hardware, operating systems, and network domain. In this complicated computing environment, Big Data secure services upon the heterogeneous infrastructure are deployed efficiently by taking the abstraction capability of Software-Defined Networking (SDN). An abstraction concept of SDN separates the control (higher) plane from the underlying system infrastructure being controlled and supervised. Segregating network's control logic from the underneath physical routers and switches that forward traffic makes the system administrators to write high-level control programs that specify the behavior of an entire network, in contrast to conventional networks, whereby administrators (if allowed to do it by the device manufacturers) must codify functionality in terms of low-level device configuration. The intelligent management of secure functions simplifies the concepts of implementation of security rules; system (re)configuration; and system evolution using SDN. The main problem of a centralized SDN solution can be reduced by using a chain of controllers and/or through the usage of more number of controllers to control at least the most important functions of a system.

The National Institute of Standards and Technology (NIST) launched a framework with a set of voluntary guidelines to help organizations make their communications and computing operations safer (NIST, 2014). This could be accomplished by verifying the system infrastructure in view of security against threats, risk assessment, and ability to respond and recover from attacks.

Following the last verification principles, Defense Advanced Research Projects Agency (DARPA) is creating a program called Mining and Understanding Software Enclaves (MUSE) to enhance the quality of the US military's software. The design of this program is to develop more robust software to work with big datasets with no errors or break down under the huge volume of information (DARPA, 2014). Additionally, main concepts like privacy and security are becoming the most important aspects of Big Data aspects that need to be addressed. As the social networks made the people to share and distribute important copyrighted digital information in a very easy way, the copyright infringement behaviors like malicious distribution, illicit copying and usage, unauthorized access, and free sharing of copyright-protected digital contents, will become a much more common phenomenon. To alleviate these problems Big Data must provide solid solutions to protect security for author's privacy and author's copyrights. In addition, it has become a common trend to share personal data through their mobile devices and computers to social networks and cloud services and misplace data and content leads to a severe (serious) impact on their own confidentiality. Hence, a secure framework to social networks is one of the most important topics in research.