Chapter 13
Big Data Security Management

Zaiyong Tang
Salem State University, USA

Youqin Pan
Salem State University, USA

ABSTRACT

Big data is a buzzword today, and security of big data is a big concern. Traditional security standards and technologies cannot scale up to deliver reliable and effective security solutions in the big data environment. This chapter covers big data security management from concepts to real-world issues. By identifying and laying out the major challenges, industry trends, legal and regulatory environments, security principles, security management frameworks, security maturity model, big data analytics in solving security problems, current research results, and future research issues, this chapter provides researchers and practitioners with a timely reference and guidance in securing big data processing, management, and applications.

INTRODUCTION

Big Data has become an entrenched part of discussions of new development in information technology, businesses, governments, markets, and societies in recent years. It has inspired noteworthy excitement about the potential opportunities that may come from the study, research, analysis, and application of big data. However, accompanying those enticing opportunities and prospective rewards, there are significant challenges and substantial risks associated with big data. One of the biggest challenges for big data is increased security risk.

Security for big data is magnified by the volume, variety, and velocity of big data. With the proliferation of the Internet and the Web, pervasive computing, mobile commerce, and large scale cloud infrastructures, today's data are coming from diverse sources and formats, at a dynamic speed, and in high volume. Traditional security measures are developed for clean, structured, static, and relatively low volume of data. Undeniably, big data presents huge challenges in maintaining the integrity, confidentiality, and availability of essential data and information.

DOI: 10.4018/978-1-4666-9840-6.ch013

At the same time as big data is gaining momentum in the networked economy, security attacks are on the rise. Security perpetrators are becoming more sophisticated, wide-spread, and organized. As businesses and other organizations are moving towards a more open, user-centric, agile, and hyper connected environment that fosters intelligence, communication, innovation, and collaboration, attackers are eager to exploit the intrinsic vulnerabilities that come with open and dynamic systems. Since big data are generated by those systems, and newly developed big data infrastructures such as NoSQL databases, cloud storage, and cloud computing have not been thoroughly scrutinized for their capability of safeguarding data resources, information security presents a formidable challenge to big data adaptation and management.

By identifying and laying out the major challenges, issues, industry trends, framework, maturity model, and fundamental principles of big data security, this chapter will provide researchers and practitioners with a timely reference and guideline in securing big data processing, management, and applications.

BACKGROUND

Big Data is generally considered to have three defining characteristics: volume, variety and velocity (Zikopoulos, et al. 2012). When at least one of the dimensions is significantly high, the data is labeled big. Traditional techniques and technologies are not sufficient to handle big data. With the enormous size, speed, and/or multiplicity, big data processing requires a set of new forms of technologies and approaches to achieve effective decision support, insight discovery, and process optimization (Lancy, 2001). Although the three V's of big data definition has wide acceptance, recently, there have been attempts to expand the dimension of big data to include value and veracity (Demchenko, et al., 2013). The value dimension of big data deals with drawing inferences and testing hypothesis, and the veracity dimension is about authenticity, accountability, availability, and trustworthiness. Some researchers (e.g., Biehn, 2013) have suggested adding value and viability to the three V's.

Although big data has been discussed for over a decade since 2000, interest in big data has only experienced significant growth in the last few years. Figure 1 shows the Google search interest for the search term "Big Data" from January 2004 to June 2014. The figure does not show actual search volume. The y axis represents search interest relative to the highest point, with the highest point being scaled to100. For more historical information about big data, the reader is referred to Press (2013), which documents the history of big data that dates back to the 1940s.

Soon after the interest in big data has gained significance, the search interest in big data security and big data privacy has also experienced remarkable growth. Figure 2 shows the Google search interest for the search term "Big Data Security" (top line) and "Big Data Privacy" (bottom line) from January 2011 to June 2014.

In order to clearly articulate, define, and build the big data ecosystem, there has been a concerted effort to develop big data architecture by both standard organizations such as National Institute of Standards and Technology (NIST) and major information technology companies such as IBM, Microsoft, and Oracle. Demchenko et al. (2013) consolidated previous models of big data architectures and proposed a Big Data Architecture Framework. This framework consists five key components of the big data ecosystem: Data Models, Big Data Management, Big Data Analytics and Tools, Big Data Infrastructure, and Big Data Security. Data model defines the type and structure of data. Big data management deals