ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

A Survey on Big Data Information Security

Prof. Rushikesh M. Shete, Prof. R. V. Chaudhari

Assistant Professor, Dept. of I.T., DMIETR, Wardha, RTM University, Nagpur, Maharashtra, India

Lecturer, Dept of CE, BDCE, Sevagram, RTM University, Nagpur, Maharashtra, India

ABSTRACT: Nowadays, attacks bring extreme threat and challenge to the information security, based on analysis of big data technique. The growing popularity and development of data mining technologies bring serious threat to the security of individual's sensitive information. In recent years challenges of Big Data IT are management of large amounts of heterogeneous information and providing its availability. In this paper we will study Big Data protection against unauthorized access and corruption (keeping its confidentiality and integrity) as well as availability maintenance form the key research priorities in this field.

KEYWORDS: Information Security; Big Data; Data Mining, Privacy-Preserving Data Mining**.**

I. INTRODUCTION

As a production of factor big data is being intensively integrated with development of various organizations. The consequent security issues concern with personal information which becomes increasingly severe. Using the technology of Big Data collecting and mining increased significantly over the past decade. This is due to the fact that a large amount of data is generated in the daily activities of the various organizations and, hence, the volume of organizational information resources grows dramatically.

The processing of Big Data by its nature exceeds the capabilities of computing resources available to the organizations and conventional information management methods. Decreasing the cost for centralized storage and appropriate handling would allow organizations to collect more information about the various aspects of their business. Information is inevitable in all kinds of entrepreneurial activities, and must be therefore protected as assets. Information security may be assured in various ways, including related policies, processes, procedures, organizational structures, software programs and hardware equipment able to eliminate many sources of safety jeopardizing such as espionage, computer fraud and deceit, sabotage, vandalism, fire or water.

Requirements for information security should cover three areas: risks to the organization, including its strategy and objectives, its potential vulnerability and the likelihood of adverse events; legislation, statutory, regulatory and contractual requirements that the organization and its contractors must comply with; principles, objectives and business requirements for processing the information, that the organization must develop in order to refrain from business failures and to support its activities. With the depending of Internet applications, social networks and internet of things generates a large amount of data, which we called big data. It makes the analysis and application of the data more complex, and difficult to manage. These data, including text, images, audio, video, Web pages, e-mail, microbe logging and other types, among them, 20% are structured data, 80% are semi-structured and unstructured data. Big data is large and complex, so it is difficult to deal with the existing database management tools or data processing application [1], [2], [3].

Why do we collect and analyze big data?

-Because we can get the benefit from it.

(1) To Acquire Knowledge. Because of big data holds a large number of unique information, big data analysis can effectively get rid of individual differences, to help people through the phenomenon, more accurately grasp the law behind the things.

(2) To Presume the Trend. Using the knowledge, we can more accurately predict the natural or social phenomenon. Google foretold the presence trend of flu around the world, through the statistics of search for flu information.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

(3) To Analyze Personality Characteristics. Commercial enterprise collect information on all aspects of customers for a long time, to analyze the user behavior law, more accurately portray the individual Characteristics, to provide users with better individualized good and services, and more precise advertising suggested. For example, e-commerce sites now use Big Data technology record customer browsing and purchasing history, to guess his interest, and recommend products for him, this may be his interest.

(4) To Discern Truth by Analyzing. In the network, data sources are distinct, type is rich, so the authenticity can't be conferred. At the same time, the spread of information on the Internet is more convenient, so the damage caused by false information on the Internet is greater. Due to the enormous amount of data in the big data environment, to a certain range, it can help discern truth by analyzing the data. Big data bring the benefits to us, but also bring the questions of data security and privacy protection, since the emergence of big data technology, a large number of security incidents have been occurred, and these incidents sounded the alarm for the society.

II. RELATED WORK

Theoretical Background

A. Risk Assessment in Information Security

Risk assessment must include assessment of the risks size (risk analysis) and its comparison with the determined criteria. This work must be repetitive owing to the possible changes in the conditions of the company operation or with regard to the possible acceptability of risks. The outcome of the risk assessment may involve: limitation of the risk occurrence; acceptance of risk and limitation of its occurrence reduction of the activities related to a given risk; delegation of the risk to another organization (insurance, suppliers). Management of the organization must define information security policy in compliance with the requirements of the organization, applicable laws and regulations. The policy has to be officially approved, published and communicated to all employees and interested parties. At planned intervals, the policy must be reviewed and communicated to all stakeholders, especially if there have been changes that might threaten its suitability, adequacy and effectiveness. As in figure 1 shows the overview of risk assessment in information security. Here the external monitoring entails the collecting the information from data sources and gives the intelligence context. The internal monitoring maintains information of awareness of state of organization network where the planning and risk assessment look's out the over information of security program also accomplished by identifying and planning ongoing information security activities that further reduce the risk [2].

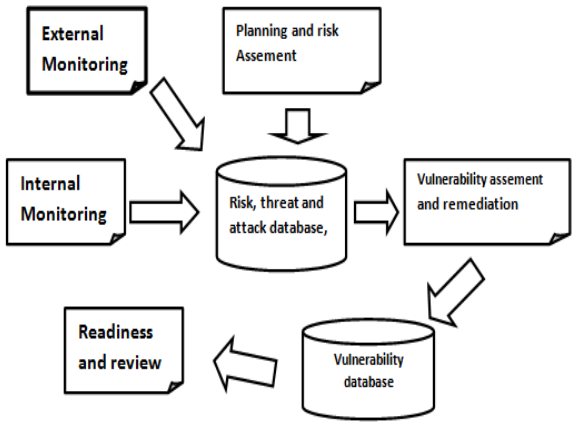

Fig. 1. Overview of risk assessment in information security

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

The primary goal of vulnerability and remediation is to report the status of vulnerability. Ensuring the proper level of management is involved. Communicating vulnerability information to owners of vulnerable systems..

B. Security Technology In Big Data Environment

For the security risks of big data, we need to address the security issues of big data from the following points: data privacy protection technology; data integrity and trusted technology; access control technology.

Access Control Technology:

Big data holds a plenty of information resources, all occupations and trades have great requirement of the data, so we must manage access rights of big data carefully. Access control is an effective means to achieve controlled sharing of data, but in big data environment, the number of users is huge, the authority is complex, and a new technology must be adopted to realize the managed sharing of information. (1) Role Mining Role-based access control (RBAC) is an access control model used widely. By appointing roles to users, roles related to permissions set, to accomplished user authorization, to simplify rights management, in order to achieve privacy protection. In the early, RBAC rights management applied "top-down" mode: According to the enterprise's position to establish roles, When applied to big data scene, the researchers begin to focus on "bottom-up" approaches, that is based on the existing "Users-Object" authorization, design algorithms automatically retrieve and optimization of roles, called role mining. In the big data scene, using role mining techniques, roles can be automatically generated based on the user's access records, efficiently provide personalized data services for mass users. It can also be used to detect potentially dangerous that user's behaviour deviates from the daily behaviour. But role mining technology are based on the exact, closed data set, when applied to big data scene, we need to solve the special problems: the dynamic changes and the quality of the data set is not higher.

How to Achieve Access Control for Big Data:

Access control is an effective method to realize data controlled sharing; it is divided into discretionary access control, mandatory access control and role-based access control. While in big data environment, it is difficult to preset the role, to realize the role and to predict the actual authority of each role. Discretionary access control is unable to meet the diversity of the permissions due to the diversity of users, mandatory access control is unable to meet the power of authority, and role-based access control is not able to effectively match the role and the corresponding permissions. Therefore, a new security access control mechanisms must be adopted to protect data in big data environment.

C. Research in the area of integrity and privacy ensuring in the analysis of Big Data

Many scientists have also worked in the field of IS properties investigation for the Big Data mining. Some authors presented the algorithms for Big Data processing ensuring privacy and integrity. The distributed Big Data mining and the use of confidential computation protocols overlap, since for both processes the computation of functions is performed by multiple users of information systems without the need for disclosure of the input data to each other. For distributed Big Data mining it is required the participants of the information exchange to perform function computations together on the basis of their protected data, preserving their integrity and confidentiality. In order to solve the issues emerged one need to develop some algorithms for secure Big Data collection, subsequent processing and analysis. To design such algorithms it is important to identify their requirements in terms of IS and to choose the methods to be used and further program implementation. In addition to the challenges of Big Data confidentiality, it is also necessary to solve the problem of ensuring the data integrity that can emerge with the substitution of data source or data itself. According to IBM, the financial losses caused by Big Data integrity violations are 3 trillion dollars per year [24]. Ensuring data integrity and transmission during the collection, delivery, acquisition, integration, categorization, correlation, analysis and further use, as well as the integrity of the intranet components themselves is critical to make the right management decisions. Thus, the objective of the research is the formulation and substantiation of the specific recommendations for developing the Big Data secure analysis algorithm based on the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

IS requirements (namely, integrity, availability and confidentiality) with respect to both the analysis itself and the analysed data (initial, intermediate and received as the result of the analysis). All these issues are particularly critical for network security monitoring systems that process large amounts of data about intranets' IS. Huge data volumes regarding the current state of the intranet and its resources and at the first glance unrelated (disparate) events taking place in it should be handled correctly and efficiently to identify the IS incidents and distinguish intranet areas, that are the most susceptible to the high risk, for their quick elimination. The data are generated from the information considered in the certain context and not only coming from the single domain controllers, proxy servers, DNS servers, information protection tools (IPTs), but also describing the current configuration of network devices, the characteristics of the network traffic, application and network service performance, activity and specific actions of individual end-users, as well as containing mail correspondence, web content, digital audio and video, business process data, internal documents and analytical data of the corporation for many years of its existence [4]. All the network traffic contains important information, transmitted in user shared environment and requiring to ensure the integrity, availability and, in some cases, confidentiality.

III. CONCLUSION

Statistics proved that there is a growing demand for the use of Big Data, it is characterized by the IS challenges unsolved so far. The paper shows that the IS issues for Big Data (namely, privacy, integrity and availability) are the subject of many scientists' research. And a security threat proposes the technology to solve the security threat, finally, study about the applications of big data in information security. Of course, with the development of big data technology, new security threat may appear, we need to find new solutions and technologies to solve it.

REFERENCES

1. Gang Zeng "Big Data and Information Security", ISSN (e): 2250 – 3005 Volume, 05 Issue, International Journal of Computational Engineering Research, Vol No 05, Issue No 06, Page No 17-21, June 2015.
2. Natalia Miloslavskaya and Aida Makhmudova "Survey of Big Data Information Security" in 4th International Conference on Future Internet of Things and Cloud Workshops, 2016.
3. FENG Deng-Guo, ZHANG Min, LI Hao. Big "Data Security and Privacy Protection [J]" Chinese Journal of Computers, Page No. 246-258, 2014.
4. D. Chaffey. Global social media research summary 2016. URL:http://www.smartinsights.com/social-mediamarketing/ social-media strategy/new-global-social-media-research/ (29.02.2016) February 2016..
5. N. Miloslavskaya., M. Senatorov, A. Tolstoy, S. Zapechnikov. Information Security Maintenance Issues for Big Security-Related Data. In Proceedings of 2014 International Conference on Future Internet of Things and Cloud FiCloud 2014. International Simposium on Big Data Research and Innovation (BigR&I). Page No27-29, August 2014.
6. WANG Yu-long, ZENG Meng-qi. "Big Data Security based on Hadoop Architecture[J]. Information Security and Communications Privacy", Page No. 83-86, 2014(7).
7. Guillermo Lafuente."The big data security challenge [J] Network Security", Page No. 12-14, 2015.(1).
8. M. Wernke, P. Skvortsov, F. Dürr, and K. Rothermel, ``A classi_cation of location privacy attacks and approaches," Pers. Ubiquitous Comput., vol. 18, no. 1, pp. 163_175, Jan. 2014.
9. M. Ghasemzadeh, B. C. M. Fung, R. Chen, and A. Awasthi, ``Anonymizing trajectory data for passenger _ow analysis," Transp. Res. C, Emerg. Technol., vol. 39, pp. 63_79, Feb. 2014.
10. Hrestak D, Picek S. "Homomorphic Encryption in the Cloud[C]" 2014 37th International Convention on Infor-mation and Communication Technology, Electronics and Micro electronics(MIPRO). 2014: 1400-1404.
11. Hu Kun, Liu Di, Liu Minghui. "Research on Security Connotation and Response Strategies for Big Data[J]. Telecommunications Science", Page No.:112-117,122,
12. R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang, ``Privacy-preserving trajectory data publishing by local. suppression," Inf. Sci., vol. 231, pp. 83_97, May 2013.

BIOGRAPHY

Mr. Rushikesh Moreshwarrao Shete is an Assistant Professor in the Information Technology Department, Datta Meghe Institute of Engineering, technology & Research, Sawangi (M), RTM University Nagpur. I received Master of Engineering (M. E.) degree in 2014 from SGBAU, Amravati, MS, India. My research interests are Computer Networks (wireless Networks), Data Mining, web 2.0.