A Security Model for Big Data Usage on Cloud Computing (SMBDCC)

Maseeh Ullah Khan[1], Abdul Wahid Khan[2]

Department of Computer Science, University of Science & Technology Bannu, KPK, Pakistan

Abstract

For the rapid growth of big data, cloud computing is a well-known approach for organization to improve the accessibility, management and processing of data on internet; but still cloud computing is not free from various security challenges.

The main objective of our research is to develop a security model for Big Data usage on Cloud Computing (SMBDCC) that will assist software vendor organization to know in advance about security challenges and their practices for using big data on cloud computing. It will assist software vendor organization to know about their status while using big data on platform of cloud computing. For the data collection of our proposed research we will use systematic literature review (SLR). Then for the verification of these SLR results, pairwise comparison and prioritization of all challenges and practices we will use questionnaire survey, Analytical Hierarchy Process (AHP) respectively, and finally the developed model would be validated by industry specialists with the help of case study. The main purpose of our proposed model is the unique contribution to support software vendor organization about security related challenges of big data usage on cloud computing.

Key words:

Big Data, Cloud Computing, Security Issues, SLR, Software Vendor.

**Corresponding author address:*
Mr. Maseeh Ullah Khan is the corresponding author of this paper. E-mail: maseeh.bangash@gmail.com.
*This work is supported by Dr. Abdul Wahid Khan. E-mail: wahidkn@gmail.com.

1. Introductions

In the recent period data is increased in very high speed and become more economical and abundant. As 2.5 quintillion bytes of data may be generated per day and increased from TB up to PB with time [1].
According to Walmart estimation he can gather up to 2.5 petabytes of client-related transactional data just within one hour [2, 3].
This large scale data include digital images, videos, social links updates, smart devices, buying business accounts, mobile GPS signals, and websites etc. [4, 5].
Data either structured, unstructured or semi structured develops big data whenever the volumes, velocity, variety, or veracity surpasses the capacity of IT system so that to stock, devour, evaluate, and process that data [2, 6].

The rapidly growth and plentiful of data not only because of the volume (size), but there are some other parameters which are the velocity, veracity, variety and the ratio of structured data (20%) and unstructured (80%) data. This data comes from public places, industry, organization, and business functions. Big data not only containing data but also includes tools, techniques, and frameworks [5, 7].
According to Apache Hadoop [1] the big data is defined as " the large amount of data of which management, capturing and processing should not be done in traditional systems with in the given time period and resources".
For handling this huge amount of data for software vendors organizations "which are basically the software publishers developing and vending software's", cloud computing are used to perform all these processes on internet. Cloud technology is the internet based technology that depends upon the distribution of shared resources, software's and information on internet instead of local server or individual devices on user demand [8].
The main objective of cloud environment is the use of maximum computational power which perform lots of commands within second. The network of cloud technology is consist of so many servers that are connected to each other for a purpose to process massive amount of data between these connected cluster. These processed data of user is keep on private storage location may be located in any other city or country in the world [9, 10].
Cloud computing mainly provides three services which are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). Example of cloud computing which we used in our daily life are Facebook, YouTube, Dropbox, and Gmail etc. which provides scalability, flexibility, agility, and simplicity by which the usage of cloud computing is quickly increased in the enterprises [9, 11, 12].
In today's era cloud computing is become so much popular. Although cloud computing has enormous benefits but still facing several security issues which are [8, 10, 12, 13]:

- Confidentiality and privacy
- Integrity
- Availability
- Trusted Third Party
- Interoperability
- Malicious Insider
- Trust

While processing huge amount of data on internet some security challenges are faced which are categorized mainly in four levels [9, 14]:

- Network level
- User authentication level
- Data level
- Generic issues

2. Literature Study

There are several security challenges for Big data on cloud environment which could affect data encryption, data storing and data sharing.

One of the most important issues is distributed nodes architecture issues in which data processing occur only in those nodes where recourses needed. As there are multiple nodes in distributed nodes so it will occur at anyplace throughout all the nodes, and it will become very difficult to find out the exact node address of processing. That's why finding of exact location of computation is very difficult to address [9, 15].

Distributed data is also an security issue at network level as data is distributed among multiple nodes and many copies of the data present for data consistency at each node so that when a file is corrupted then it is recovered from its copies but in cloud computing finding of the exact address for a particular data file that where it is stored is so much tedious and hard [9, 15].

In cloud computing for data communication between nodes, wireless or wire networks are used. So any one can easily beat or alter the inter node communication for making infringement in the system [9, 15].

Another issue is that for better performance and efficacy cloud computing like Hadoop, stock data without encryption. So hacker can easily access to that critical data and no one can stop his access to secret data [9, 15, 16].

As there are multiple nodes connected in cloud computing and they all have an unrestrained access to the data. Due to which malevolent node can either steal or change the important data. So this free administrative access of every node to data is an security issue in cloud computing [9, 15].

Authentication of users and nodes is also a security issue in cloud computing because in cloud computing nodes are linked with clusters to improve multi-processing. But having no authenticity malevolent nodes can get access to the clusters and steal important data or interrupt the operations [9, 15, 16].

For deletion, modification and storing of data in cloud environment without logging no action is recorded. Without loggings finding of any malicious user if has broken the connected bunch and change the data is not an easy job [9, 15].

In cloud computing as there are numerous technologies having lot of interrelating compound components that includes database, computing power, network and many others. Due to this extensive use of technologies and interrelation of components whenever there will a trivial security problem arise in single part could affect the overall environment. Therefore keeping security in cloud computing is a challenging task [9, 15].

In cloud environment use of traditional security instruments is also an issue as designing of these security instruments are only suitable for traditional computing environment of which accessibility and expendability was small than that of cloud environment. These traditional security instruments could not directly applicable in cloud environment which are developed earlier. These traditional instruments cannot scale as compared to cloud scale [9].

Cloud security association also discusses the main security issues for big data in cloud environment which are namely [8, 14]:

- “Secure computation in distributed programming frameworks”
- “Security best practices for non-relational data bases”
- “Secure data storage and transactions logs”
- “End-point input validation/filtering”
- “Real-time security monitoring”
- “Scalable and composable privacy-preserving data mining and analytics”
- “Cryptographically enforced data centric security”
- “Granular access control”
- “Granular audits”
- “Data source”

3. Problem Statement

We have found through literature survey that different Researchers have identified many security challenges regarding Big Data usage on Cloud Computing but the study only based on single dimension of the issue where every particular researcher point out a single issue. However, with the rapid growth of data on cloud technology there is a need of such a model which not only discusses the issues but also to provide proper solutions for these issues that will provide road map to software vendor organization in context of big data usage issues on cloud computing. After a detailed study of literature review we find out that still there is no such model which based on these challenges facing by Big Data users on Cloud Computing using the concept of SLR. In light of the above facts and figures we have planned to develop a model that will assist software vendor organization to know about the critical challenges and their practices when using a concept of big data on the platform of cloud environment.

4. Significance

We have find out through literature review that there are numerous issues of big data usage on cloud computing platform. These issues are discussed by many researchers where they have defined individually and we have not found any such model using SLR to support software vendor organizations on usage of big data on the platform of cloud computing. Our proposed security model for Big Data usage on Cloud Computing (SMBDCC) will provide solution for security related issues. Our proposed model is unique contribution to support software vendor organization about security related issues of big data on cloud environment.

5. Research Objectives

The core objective of our research is to develop a security model for Big Data usage on Cloud Computing (SMBDCC) to identify and solve security related challenges for software vendor organization. Our proposed model consists of following steps which are discussed as under:

- Conducting Systematic Literature Review (SLR) to identify big data security issues facing by software vendors organization in cloud computing.
- Conducting Empirical Study (questionnaire survey) to validate the SLR conclusions and finding of issues/practices other than already been discussed.
- Conducting of Analytical Hierarchy Process (AHP) for pairwise comparison and prioritization of these challenges and practices for our proposed model.
- Conducting Case Study for our proposed model for Big Data usage on Cloud Computing (SMBDCC).
- Conducting of Focal Group Discussion (FGD) about identification of security challenges, validation and verification of our proposed model.

6. Research Question's

We have the following three research questions which are the core objective of our research:

RQ1: What are the main security issues / challenges faced by the software vendor organization to secure the Big Data on Cloud Computing?
RQ2: What are the existed practices / solutions as defined in the background review used to avoid the security issues faced by the software vendor organization?
RQ3: What are the practices / solutions of real world used to avoid the Big Data Security issues /challenges faced by the software vendor organization on cloud computing apart from identified?

7. Research Methodology

Research methodology of our proposed research is consisting of four steps. First of all finding of main issues and its practices which are faced by software vendor organization for big data on cloud computing using systematic literature review (SLR). Secondly conducting of Empirical Study (questionnaire survey) to validate the SLR conclusions and also findings of new issues/practices other than already been discussed. Thirdly conducting of Analytical Hierarchy Process and Case Study for our proposed model for Big Data usage on Cloud Computing (SMBDCC). And finally conducting of Focal Group Discussion (FGD) about identification of security challenges, validation and verification of our proposed model.

8. Research Plan

Research plan of our proposed model is that first of all we will to identify all the security challenges through systematic literature review (SLR) which are faced by software vendor's organization. Secondly, Questionnaire survey will be conducted for the validation of these identified issues which are finding through SLR and also to identify new ones apart from the identified ones. Thirdly, on the bases of SLR and Questionnaire survey findings we will finalized our proposed model and conduction of Analytical Hierarchy Process (AHP) for our proposed model so that to prioritized and compared pairwise each challenge and practice. Finally, case study will be conducted in software industry to get feedback from the experts about the reliability of our proposed model.

9. Model Development

There are three phases which are elaborated for the development of SMBDCC as described in Figure 1. These three steps are discussed below:

- Step 1 will identify all those challenges, barriers or complications which affect big data security on cloud computing. These challenges will be identified through Systematic Literature Review (SLR).
- Step 2 will identify all those practices that are adopted by software vendor organizations in order to avoid different challenges of big data security on cloud computing. These practices will find out with help of SLR and Empirical study.
- Step 3 is concerned for the evaluation of security model for Big Data usage on Cloud Computing (SMBDCC). After successful designing of SMBDCC case study will be conducted for its evaluation. And at last Supervisor input, Analytical Hierarchy Process (AHP) and Systematic Literature Review will also contribute in evaluation of this model.

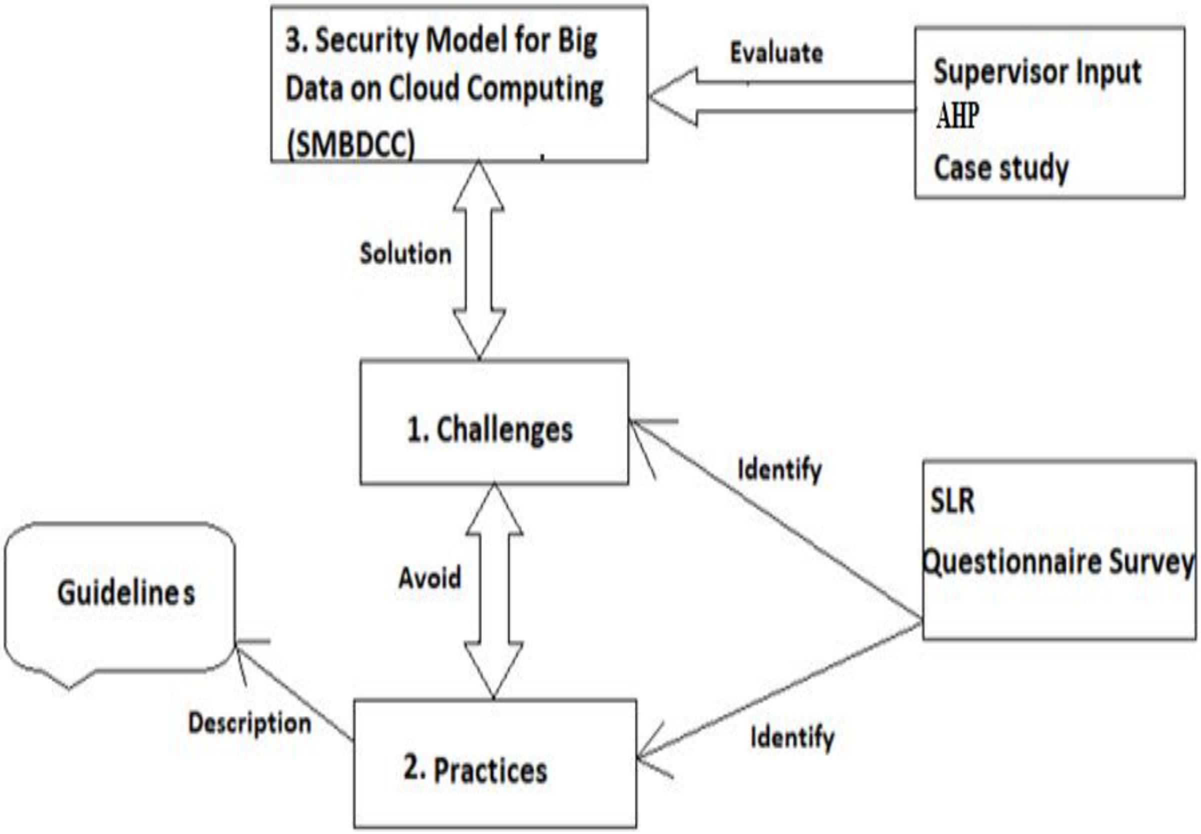

Figure 1. A Security Model for Big Data on Cloud Computing.

10. Acknowledgement

We are very much thankful to the University of Science and Technology, Bannu, for its endless support and contribution in our research.

11. References

1. Chen, M., et al., Big data: related technologies, challenges and future prospects. 2014.
2. Kruse, C.S., et al., Challenges and opportunities of big data in health care: a systematic review. JMIR medical informatics, 2016. 4(4): p. e38.
3. McAfee, A., et al., Big data: the management revolution. Harvard business review, 2012. 90(10): p. 60-68.
4. Varsha Jambunathan1, S.V., A Review on Big Data Challenges and Opportunities. International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) November 2016 Volume V(Issue XI): p. 4.
5. A.Saeed, E.S.A.A.a.R., A Survey of Big Data Cloud Computing Security International Journal of Computer Science and Software Engineering (IJCSSE), 2014. 3(1): p. 7.
6. S.Harini, K.J.a.K.J., A review of big data computing and cloud International Journal of Pure and Applied Mathematics 2018. 118 (18): p. 8.
7. Swamil Singh1, V.M., Dr. Sanjay Srivastava3 The Big Data analytics with Hadoop: Review International Journal for Research in Applied Science & Engineering Technology (IJRASET) 2016 4(III): p. 5.
8. Tiwari, P.K. and B. Mishra, Cloud computing security issues, challenges and solution. International journal of emerging technology and advanced engineering, 2012. 2(8): p. 306-310.
9. Inukollu, V.N., S. Arsi, and S.R. Ravuri, Security issues associated with big data in cloud computing. International Journal of Network Security & Its Applications, 2014. 6(3): p. 45.
10. Sabir, S., Security Issues in Cloud Computing and their Solutions: A Review (IJACSA) International Journal of Advanced Computer Science and Applications, 2018 9: p. 4.
11. khan, P.S.a.R., A Review Paper on Cloud Computing International Journals of Advanced Research in Computer Science and Software Engineering, 2018. 8(6): p. 4.
12. Laure, A.G.a.E., BIG DATA SECURITY AND PRIVACY ISSUES IN THE CLOUD. International Journal of Network Security & Its Applications (IJNSA), 2016 8(1): p. 20.
13. Tripathi, P. and M. Suaib, Security Issues On Cloud Computing. International Journal of Engineering Technology, Management and Applied Sciences, 2014. 2(6).
14. Kumar, S.S., J. Lambodar, and S. Santhosh, Big Data Security issues and challenges in Cloud Computing Environment. International journal of Innovation in Engineering and Technology (IJIET), 2015. 6: p. 297-306.
15. Securosis, L., Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments. 2012.
16. Neves, P.C., et al. Big Data in Cloud Computing: Features and Issues. in IoTBD. 2016.