

Survey on Big Data Security in Cloud Computing Environments & Future Research Challenges

G. Kishore Kumar¹ Dr. M. Gobi²

¹Research Scholar ²Assistant Professor

^{1,2}Department of Computer Science

^{1,2}Chikkanna Government Arts College, Tiruppur India

Abstract— Privacy and Security issues are extended by velocity, volume and variety of big data such as high-volume inter-cloud migration, mixture of data sources and formats, flowing nature of data acquirement, and large-scale cloud infrastructures. In all the areas especially for marketing and research, most of the industries uses big data, but they do not have the essential assets particularly from the point of security perspective. If any security breach occurs to big data, it would result in even further serious legal repercussions and reputational impairment than the current situation. Hence, traditional security mechanism for securing data are inadequate. There are various security challenges such as Reliability of devices collecting data, Source validation and filtering of data, Application software security, Access control and authentication, Interoperability of devices, Distributed systems security (DDoS attack) etc. In this paper, we highlight the data security and privacy challenges proposed by various sources and proposed Security Architecture for big data in the cloud environments. For making big data secure, techniques such as encryption using ECC/HECC is endorsed.

Key words: Big Data, Challenges, Security, Privacy, Cloud

I. INTRODUCTION

Cloud computing is denoted as “cloud” which means the supply of on-demand computing resources to store, manage, and process data over the internet on pay-for-use basis. It is an Internet-based computing type, which is offered on demand to computer/other devices that needs data and shared processing resources. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in either privately owned, or third-party data centers that may be located far from the user—ranging in distance from across a city to across the world. [As per US National Institute of Standards and Technology (NIST)].

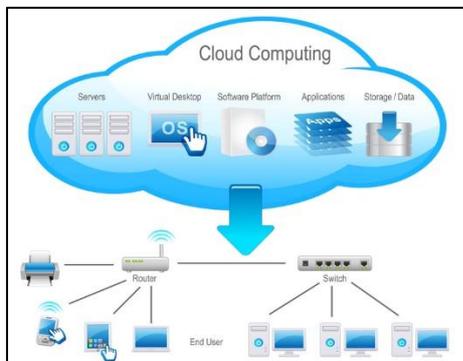


Fig. 1: Cloud Computing

A. Big Data

Big Data stands for, massive volumes of structured and unstructured data that are extensively large; hence, it is very

difficult to process those using traditional databases and software technologies. Moreover, the companies who had to query in this lightly structured very large distributed data.

Traditionally, a number of companies like large-scale Internet search, advertising, and social networking companies etc. lead the way to Big Data hardware and software inventions. For example, there is an amazing achievement in computer science & engineering wherein Google evaluates the content, links and clicks on 1.5 trillion page views per day of ALEXA.COM [An Amazon.com Company] and delivering the search results along with personalized advertising in milliseconds of time.

Similar like Google, Oracle Corp., Yahoo and others as well contributed their technologies by providing Big Data work to the open source community, broader commercial and public sector interest took up the challenge. The big data is seen differently in the market as opposed to innovators such as accumulating the new data to the existing analytical or operational systems.

The need of data secure can be in Networks, memory, Access through tools, Data Reservoir, Data Warehouse and discovery lab.



Fig. 2: Big Data

Big Data defined by the four “V”s: Veracity, Velocity, Variety and Volume. These become a realistic test for determining whether to add Big Data to information architecture.

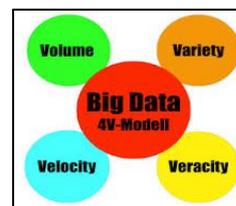


Fig. 3: Four V’s Model in Big Data

We all have heard of these V’s only yet, but Indrapal Bhandar [Chief Data Officer at Express Scripts provided in his presentation at the Big Data Innovation Summit in Boston] states that there are additional “V”s are to be concerned with IT, business and data scientists, remarkably Veracity. Other new big data V’s getting eminent are: validity and volatility.

1) Volume

Volume denotes massive volumes of data and the volume of it is exponentially growing day to day from petabytes to Zetabytes. Now this data is generated by machines, networks and human-system interactions etc as opposed to traditional created data by employees. The volume of data does not have issues like Veracity.

2) Variety

Variety refers to many types of data including unstructured, standard, raw, semi-structured and structured data, produced from various categories. This type of data cannot be handled by traditional systems, and need to be dealt with big challenges.

3) Velocity

Velocity denotes the pace at which data flows-in from various sources like business machines, networks, processes and human-system interactions like mobile devices, social media sites etc. Big data will be able to handle the incoming and outgoing data speedily and this is massive and continuous.

4) Veracity

Veracity refers to the noise, abnormality and biases in data. This means whatever be the data being stored and mined has some meaning and helping in the problem being analyzed. Veracity in data analysis is the biggest challenge when compared to volume and velocity. Moreover, various processes are required to keep away "dirty data" from accumulating in the systems.

5) Validity

Validity denotes that whether the data correct and accurate for the proposed use as valid data is the key for making the precise decisions.

6) Volatility

Volatility refers to how long the data is valid and how long it should be stored. In the current real time data world, we need to define at what point of time the data is no longer related to the required current analysis.

B. Big Data & Cloud

Cloud computing plays a vital role for Big Data; not only due to the provision of tools and infrastructure, but also because it is a business model that can be followed by Big Data analytics (e.g. Analytics as a Service (AaaS) or Big Data as a Service (BDaaS)). However, BDaaS/ AaaS brings numerous challenges as the customer and provider's staff are much more involved in this circle than in traditional Cloud providers offering software/infrastructure/platform/ as a service.

C. Big Data Security

Big Data have need of the same sort of security practices and principles. Enterprise security management practice consists of various factors like authorize resources, centralize access and manage through extensive audit practices. As various factors like Big Data technologies, data sources and user requirements addition to these practices affects the existing security management and this needs to be aligned with policies and practices already established to avoid duplicate/repetitive implementations, and manage centrally across the environments.

We would require integrated approach across all the areas as Big Data involves Oracle and non-Oracle databases. Non-Oracle database include HDFS, Hadoop, MapReduce,

Hive and Oozie. The approaches would be monitoring the database traffic, detection and blocking of threats etc.

II. RECENT RESEARCH PAPERS PUBLISHED

Below given are the research papers published related to big data security in Cloud computing:

- Marcos D. Assunção, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya, "Big Data computing and clouds: Trends and future directions", J. Parallel Distrib. Comput. 79–80 (2015) 3–15 talks about Survey of solutions for carrying out analytics and Big Data on Clouds, Identification of gaps in technology for Cloud-based analytics and Recommendations of research directions for Cloud-based analytics and Big Data.
- Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri, "Security Issues Associated With Big Data In Cloud Computing", International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014 DOI: 10.5121/ijnsa.2014.6304 45 talks about various security measures which would improve the security of cloud computing environment and various solutions which collectively will make the environment secure.
- S. Justin Samuel, Koundinya RVP, Kotha Sashidhar and C.R. Bharathi, "A Survey on Big Data And Its Research Challenges", VOL. 10, NO. 8, MAY 2015 ISSN 1819-6608 ARPJ Journal of Engineering and Applied Sciences talks about a comprehensive review of the big data state of the art, conceptual explorations, major benefits, and research challenging aspects. In addition to that, several future directions for big data research are highlighted.
- Divyakant Agrawal, Sudipto Das, Amr El Abbadi, "Big Data and Cloud Computing: New Wine or just New Bottles?", Proceedings of the VLDB Endowment, Vol. 3, No. 2 talks about design challenges which application and system designers face in developing and deploying new applications and systems, and expand on some of the major challenges that need to be addressed to ensure the smooth transition of applications from traditional enterprise infrastructures to the next generation of cloud infrastructures.
- Mingmin Chi, Antonio Plaza, Jon Atli Benediktsson, Zhongyi Sun, Jinsheng Shen, and Yangyong Zhu, "Big Data for Remote Sensing: Challenges and Opportunities", 2015 talks about the most challenging issues in managing, processing, and efficient exploitation of big data for remote sensing problems. In addition, two case studies discussing the use of big data in remote sensing are demonstrated.
- Changqing Ji, Yu Li, Wenming Qiu, Uchechukwu Awada, Keqiu Li, "Big Data Processing in Cloud Computing Environments", 2012 International Symposium on Pervasive Systems, Algorithms and Networks talks about several big data processing technics from system and application aspects. Presented the key issues of big data processing, including cloud-computing platform, cloud architecture, cloud database and data storage scheme. Following the MapReduce parallel processing framework.

- Santoshi Tsuchiya, Hoshinori Sakamoto, Yuichi Tsuchimoto & Vivian Lee, "Big Data Processing in Cloud Environments", FUJITSU Sci. Tech. J. Vol.48, No.2(April 2012) talks about
- Cloud Security Alliance, "Big Data Analytics for Security Intelligence", September 2013 talks about Big Data security analytics outlines some of the fundamental differences from traditional analytics and highlights possible research directions in Big Data security.
- Cloud Security Alliance, "Top Ten Big Data Security and Privacy Challenges", November 2012 talks about top ten security challenges based on surveys & interviews with CSA members.
- Kalyani Shirudkar, Dilip Motwani, "Big-Data Security", Volume 5, Issue 3, March 2015 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering talks about pivotal choices for big data to leverage this mature security and privacy technology, while identifying remaining research challenges.
- White paper developed by leading researchers across the United States, "Challenges and Opportunities with Big Data" talks about various challenges & research opportunities associated in big data.
- L.W.M. Wienhofena, B.M. Mathisen, D. Roman, "Empirical Big Data Research: A Systematic Literature Mapping", 10 Sep 2015 talks about systematic mapping method with which we mapped the collected research according to the labels Variety, Volume and Velocity. In addition, they addressed the application areas of Big Data.
- Duncan Waga1, Kefa Rabah, "Environmental Conditions' Big Data Management and Cloud Computing Analytics for Sustainable Agriculture", World Journal of Computer Application and Technology 2(3): 73-81, 2014 talks about environmental conditions' data like rainfall, winds, temperature etc and the use of particular cloud computing analytical tool to get some meaningful information from it which can be utilized by farmers for strategic and successful Agriculture. Previous similar studies are discussed and recommendations given.
- White Paper by Fran Howarth, "Harnessing big data for security", February 2013 talks about some of the challenges of harnessing big data security and outlines some of the key considerations and capabilities that organizations should consider when selecting a system that can handle the whole gamut of needs in a unified manner that is simple to integrate and manage.
- Dr. Jangala. Sasi Kiran, M.Sravanthi, K.Preethi, M.Anusha, "Recent Issues and Challenges on Big Data in Cloud Computing", IJCST Vol. 6, Issue 2, April - June 2015 talks about long term challenges that require research and new paradigms. Analyzing the issues and challenges comes first as we begin a collaborative research program into methodologies for big data analysis and design.
- Zan Mo, Yanfei Li, "Research of Big Data Based on the Views of Technology and Application", American Journal of Industrial and Business Management, 2015, 5, 192-197 talks about the development of technologies in big data analysis and storage and analyses the trends and different values in commercial applications, manufacturing, biomedical science and other applications.
- Raghavendra Kune, Pramod Kumar Konugurthi, Arun Agarwal, Raghavendra Rao Chillarige and Rajkumar Buyya, "The anatomy of big data computing", Software: Practice And Experience Softw. Pract. Exper. 2016; 46:79-105 talks about the evolution of big data computing, differences between traditional data warehousing and big data, taxonomy of big data computing and underpinning technologies, integrated platform of big data and clouds known as big data clouds, layered architecture and components of big data cloud, and finally open-technical challenges and future directions.
- Raghav Toshniwal, Kanishka Ghosh Dastidar, Asoke Nath, "Big Data Security Issues and Challenges", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 2, Volume 2 (February 2015) talks about essential concepts of Big Data and processes involved in data processing & analysis on the security parts of Big Data and propose a new system for Big Data Security and as a final point present the future scope of Big Data.
- Ali Gholami and Erwin Laure, "Big Data Security And Privacy Issues In The Cloud", International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.1, January 2016 DOI : 10.5121/ijnsa.2016.8104 59 talks about the categorize the existing research according to the cloud reference architecture orchestration, resource control, physical resource, and cloud service management layers, in addition to reviewing the recent developments for enhancing the Apache Hadoop security as one of the most deployed big data infrastructures. We also outline the frontier research on privacy-preserving data-intensive applications in cloud computing such as privacy threat modeling and privacy enhancing solutions.
- Jaseena K.U. and Julie M. David, "Issues, Challenges, And Solutions: Big Data Mining" talks about the literature review about the Big data Mining and the issues and challenges with emphasis on the distinguished features of Big Data. Also it deliberates some more methods to deal with big data.
- Julio Moreno, Manuel A. Serrano and Eduardo Fernández-Medina, "Main Issues in Big Data Security", future internet September 2016 talks about results obtained after applying a systematic mapping study to security in the Big Data ecosystem. It is almost impossible to carry out detailed research into the entire topic of security, and the outcome of this research is, therefore, a big picture of the main problems related to security in a Big Data system, along with the principal solutions to them proposed by the research community.
- Elmustafa Sayed Ali Ahmed and Rashid A.Saeed, "A Survey of Big Data Cloud Computing Security", International Journal of Computer Science and Software Engineering (IJCSSE), Volume 3, Issue 1, December 2014 talks about a Survey of big data with clouds computing security and the mechanisms that used to protect and secure also have a privacy for big data with an available clouds.

- Zhang Hongjun, Hao Wenning, He Dengchao1, Mao Yuxing, "Survey of Research on Information Security in Big Data", CSBC 2014 talks about the characteristics of big data information security, and focuses on conclusion of security problems under the big data field and the inspirations to the development of information security technology. Finally, this paper outlooks the future and trend of big data information security.
- Elisa Bertino, "Data Security – Challenges and Research Opportunities", SDM 2013, LNCS 8425, pp. 9–13, 2014 talks about open issues, such as data protection from insider threat and how to reconcile security and privacy, and outline research directions.

III. RESEARCH TREND IN BIG DATA

Below given are the research trends in big data:

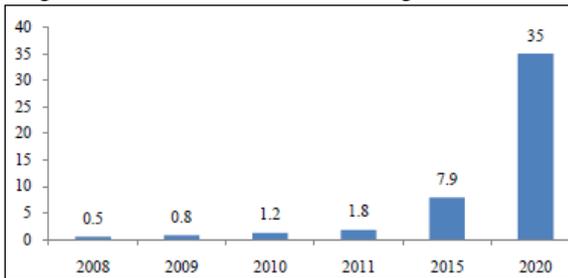


Fig. 4: The forecast of global data growth (unit: ZB)^[17]

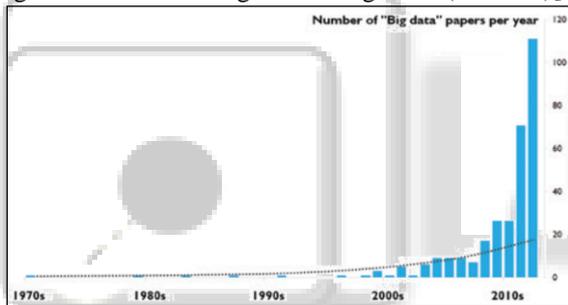


Fig. 5: No. of Big Data Papers^[26]

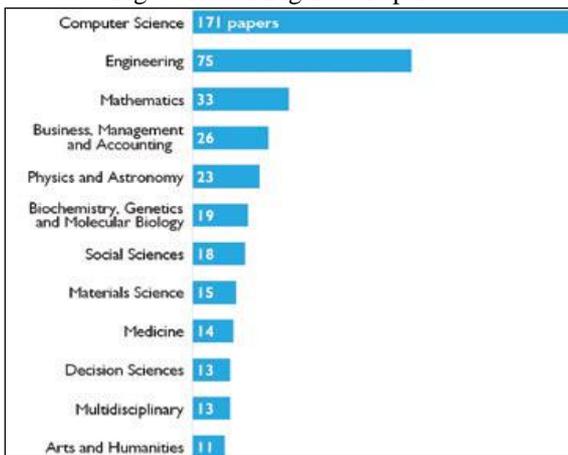


Fig. 6: Subject areas researching Big Data^[26]

IV. FUTURE RESEARCH DIRECTIONS

Cloud computing develops various security issues due to the integration of several technologies such as resource scheduling, memory management, networks, operating systems, virtualization, load balancing, databases and transaction management. In addition to this, the association of Big Data as well fetches more security threats. Data

security not only encrypting the data, but also ensuring appropriate policy implementations. In addition, the resource allocation and memory management algorithms also have to be highly secured. In our research, we are going to present various security measures, to improve/enhance the security of cloud computing environment in any of the identified six major areas in big data:

- Security
- Applied ontology
- Mobility
- Storage and Transport
- Accessibility
- Inconsistencies

ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
CSA	Cloud Security Alliance
SQL	Structured Query Language
NoSQL	Not Only SQL
HDFS	Hadoop Distributed File System
GFS	Google File System

REFERENCES

- [1] Marcos D. Assunção, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S. Netto, Rajkumar Buyya, "Big Data computing and clouds: Trends and future directions", *J. Parallel Distrib. Comput.* 79–80 (2015) 3–15
- [2] Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri, "Security Issues Associated With Big Data In Cloud Computing", *International Journal of Network Security & Its Applications (IJNSA)*, Vol.6, No.3, May 2014 DOI : 10.5121/ijnsa.2014.6304 45
- [3] S. Justin Samuel, Koundinya RVP, Kotha Sashidhar and C.R. Bharathi, "A Survey On Big Data And Its Research Challenges", *VOL. 10, NO. 8, MAY 2015 ISSN 1819-6608 ARPJ Journal of Engineering and Applied Sciences*
- [4] "An Enterprise Architect's Guide to Big Data: Reference Architecture Overview" Oracle Enterprise Architecture White Paper | March 2016
- [5] Divyakant Agrawal, Sudipto Das, Amr El Abbadi, "Big Data and Cloud Computing: New Wine or just New Bottles?", *Proceedings of the VLDB Endowment*, Vol. 3, No. 2
- [6] Changqing Ji, Yu Li, Wenming Qiu, Uchechukwu Awada, Keqiu Li, "Big Data Processing in Cloud Computing Environments", 2012 International Symposium on Pervasive Systems, Algorithms and Networks
- [7] Santoshi Tsuchiya, Hoshinori Sakamoto, Yuichi Tsuchimoto & Vivian Lee, "Big Data Processing in Cloud Environments", *FUJITSU Sci. Tech. J.* Vol.48, No.2(April 2012)
- [8] Cloud Security Alliance, "Big Data Analytics for Security Intelligence", September 2013
- [9] Cloud Security Alliance, "Top Ten Big Data Security and Privacy Challenges", November 2012
- [10] Kalyani Shirudkar, Dilip Motwani, "Big-Data Security", *Volume 5, Issue 3, March 2015 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering*

- [11] white paper developed by leading researchers across the United States, "Challenges and Opportunities with Big Data"
- [12] L.W.M. Wienhofena, B.M. Mathisen, D. Roman, "Empirical Big Data Research: A Systematic Literature Mapping", arXiv:1509.03045v1 [cs.DL] 10 Sep 2015
- [13] Duncan Wagal, Kefa Rabah, "Environmental Conditions' Big Data Management and Cloud Computing Analytics for Sustainable Agriculture", World Journal of Computer Application and Technology 2(3): 73-81, 2014
- [14] White Paper by Fran Howarth, "Harnessing big data for security", February 2013
- [15] Oracle Enterprise Architecture White Paper, "An Enterprise Architect's Guide to Big Data : Reference Architecture Overview", MARCH 2016
- [16] Dr. Jangala. Sasi Kiran, M.Sravanthi, K.Preethi, M.Anusha, "Recent Issues and Challenges on Big Data in Cloud Computing", IJCST Vol. 6, Issue 2, April - June 2015
- [17] Zan Mo, Yanfei Li, "Research of Big Data Based on the Views of Technology and Application", American Journal of Industrial and Business Management, 2015, 5, 192-197
- [18] Raghavendra Kune, Pramod Kumar Konugurthi, Arun Agarwal, Raghavendra Rao Chillarige and Rajkumar Buyya, "The anatomy of big data computing", Software: Practice and Experience Softw. Pract. Exper. 2016; 46:79-105
- [19] Raghav Toshniwal, Kanishka Ghosh Dastidar, Asoke Nath, "Big Data Security Issues and Challenges", International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Issue 2, Volume 2 (February 2015)
- [20] Ali Gholami and Erwin Laure, "Big Data Security and Privacy Issues in the Cloud", International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.1, January 2016 DOI : 10.5121/ijnsa.2016.8104 59
- [21] Jaseena K.U. and Julie M. David, "ISSUES, Challenges, And Solutions: Big Data Mining"
- [22] Julio Moreno, Manuel A. Serrano and Eduardo Fernández-Medina, "Main Issues in Big Data Security", future internet September 2016
- [23] Elmustafa Sayed Ali Ahmed and Rashid A.Saeed, "A Survey of Big Data Cloud Computing Security", International Journal of Computer Science and Software Engineering (IJCSSE), Volume 3, Issue 1, December 2014
- [24] Zhang Hongjun, Hao Wenning, He Dengchao, Mao Yuxing, "Survey of Research on Information Security in Big Data", CSBC 2014
- [25] Elisa Bertino, "Data Security – Challenges and Research Opportunities", SDM 2013, LNCS 8425, pp. 9–13, 2014
- [26] Research Trends Issue 30 September 2012 - Special Issue on Big Data