STUDY AND ANALYSES OF SECURITY LEVELS IN BIG DATA AND CLOUD COMPUTING

K.P.Maheswari[1] , P.Ramya [2],S.Nirmala Devi[3]

[1,2]Department of IT & Networking, [3]Department of Computer Science,

Subbalakshmi Lakshmipathy College of Science (India)

ABSTRACT

Cloud computing, otherwise termed as composition of various technologies such as Networks, Databases, Operating systems, Virtualization, Resource scheduling, Transaction management, Load balancing, Concurrency control and Memory management. The data in the cloud is stored by the service provider, which is capable and having a technique to protect data to ensure security. But, the increased volume, variety and various data types made the scenario completely changed by the concept of Big Data. It can be considered as a magic wound, where it collect, store, analyze, process and visualize huge amount of data. It provides a ways to improve data service delivery especially in the cloud. Cloud computing on the other hand helps in tackling the issues of storage and data services. It has various security issues in terms of users, data, connectivity, storage, etc. These security issues are broadly categorized in to four levels as Network Level, Authentication level, Data level and Generic type. In this paper, the comparative study is made on the security levels ,which are considered as issues common to Cloud computing and Big Data and the conclusion is made based on the impact of those levels over security.

Keywords*: Authentication, Big Data, Cloud computing, Security levels, Security issues*

I. INTRODUCTION

Big data presents a tremendous opportunity for enterprises across industries. By tapping into new volumes and varieties of data, scientists, executives, marketers, and a range of others can start making more informed plans and decisions. Without the right security big data can mean big problems. Such security challenges are focused in this paper by analyzing the processes, which is one level up towards the solution for security related issues.

II. CLOUD COMPUTING

The word “Cloud” means “The Internet” Cloud Computing means a type of computing in which services are delivered through the internet. It is a technology which depends on sharing of computing resources than having local servers or personal devices to handle the applications.

Cloud computing consists of a front end and back end .The front end includes the user’s computer and software required to access the cloud network. The back end consists of various computers, servers and database systems that create the cloud.

The user can access applications in the cloud network from anywhere by connecting to the Cloud using the Internet. The only thing that must be done at the user's end is to run the cloud interface software to connect to the cloud. Some of the real time applications which use Cloud Computing are Gmail, Google Calendar, Google Docs etc.

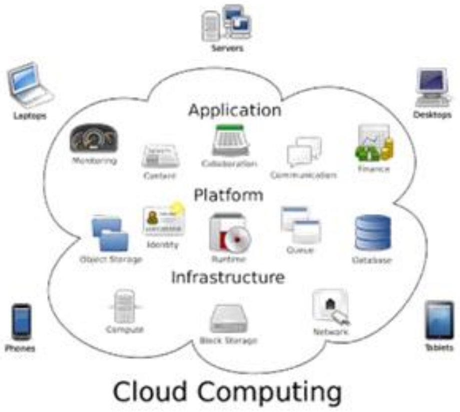


Fig.1 Cloud Computing

2.1 CLOUD SERVICES

Cloud computing can be defined based on the services offered and deployment models. According to the different types of services offered, cloud computing can be consist of three layers.Infrastructure as a Service (IaaS) is the lowest layer that provides basic infrastructure support service. Platform as a Service (PaaS) layer is the middle layer, which offers platform oriented services, and providing the environment for hosting user’s applications. Software as aService (SaaS) is the topmost layer which features a complete application offered as service on demand.

2.3 SECURITY IN CLOUD COMPUTING

Cloud computing security is the set of control-based technologies and policies designed to adhere to regulatory compliance rules and protect information, data applications and infrastructure associated with cloud computing use.As cloud computing continues to evolve, it has become increasingly important to ensure the security of virtual machines in cloud-based environments. Cloud computing security addresses both physical and logical security issues across all the different service models of software, platform and infrastructure. It also addresses how these services are delivered (public, private or hybrid delivery model).Some of the risks are linked to weak cloud security measures of the services, such as storing data without controls such as encryption, or lack of multi-factor authentication to access the service[1].An Efficient cloud security architecture should recognize the issues that will arise with security management. The security management addresses these issues with security controls such as Deterrent controls, Preventive controls, Detective controls, and Corrective controls.

III. BIG DATA

Big Data, from the name it reflects its nature,which deals with the massive volumes of structured and unstructured data that are so large that it is very difficult to process this data using traditional databases and software technologie[2].The three main properties of Big Data are, Volume: Many factors contribute towards increasing Volume streaming data and data collected from sensors etc.,

a) Variety: Today data comes in all types of formats emails, video, audio, transactions etc.,

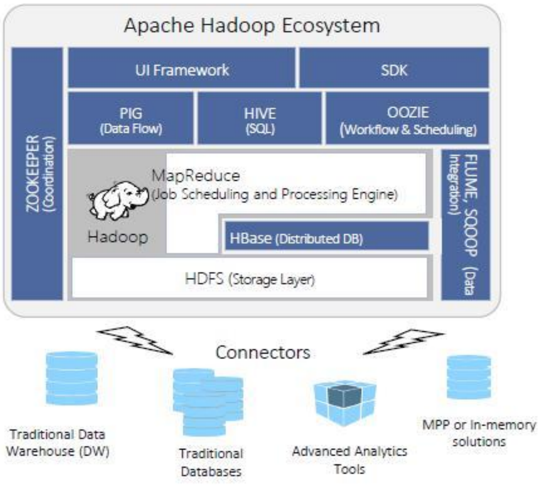
b) Velocity: This means how fast the data is being produced and how fast the data needs to be processed to meet the demand. d)Variability: This goes along with velocity, and it has to do with how inconsistent the flow of data can be with respect to time and far it can go.

e)Complexity: the complexity of the data must be considered especially when we have multiple source of data. The data must be rearranged in such a format that will be suitable for processing.

Fig.2 Big Data

3.1 Hadoop

In a Big data ,the processing of large sets of data in a distributed computing environment is supported by Hadoop, which is a free, Java-based programming framework. Hadoop cluster uses a Master/Slave structure, which can process a large set of data across a cluster of servers and applications can be run on systems with thousands of nodes involving thousands of terabytes. For handling those network from the failure it uses Distributed file system, which supports rapid data transfer rates and allows the system to continue its normal operation and the risk of system failure gets reduced. Hadoop provides computing solution that is scalable, cost effective, flexible and fault tolerant. Hadoop Framework is used by popular companies like Google, Yahoo, Amazon and IBM etc., to support their applications involving huge amounts of data. Hadoop has two main sub tasks – Map Reduce and Hadoop Distributed File System (HDFS)[3].



3.2 Map Reduce

Hadoop Map Reduce is a part of Hadoop framework. It process large amounts of data in parallel on clusters of commodity hardware resources used to write applications that process large in reliable and fault tolerant manner .It first divides the data into individual chunks which are processed by Map jobs in parallel. The outputs of the maps sorted by the framework are then input to the reduce tasks. Generally the input and the output of the job are both stored in a file-system. Scheduling, Monitoring and re-executing failed tasks are taken care by the framework.

3.3 Hadoop Distributed File System (HDFS)

HDFS is a file system used for data storage and it covers all the nodes in a Hadoop cluster. It links together file systems on local nodes to make it into one large file system. HDFS improves reliability by replicating data across multiple sources to overcome node failures.

3.4 Cloud Computing In Big Data

The Big Data emerges from the growth of cloud computing and cloud data storage. Cloud computing uses a standardized technologies, which is deviated from the traditional one.In other hand we can say that Big Data is the optimal choice for cloud computing, since it needs immense computing power and storage.

3.5. Need For Security In Big Data

Nowadays many companies focus on the technologies used for storage of their huge amount of data. The main purpose is for security. Once their data are not secured it results in the black mark for their business. When Big data is used and if a security is violated it would result in even more serious legal repercussion and reputational damage than at present. The information classification are even more critical. Techniques such as encryption, logging, honeypot detection must be necessary for bigdata security. The challenge of detecting and preventing advanced threats and malicious intruders must be solved using big data style analysis. Apart from the security, privacy is also important for organization and so they should focus on the privacy issues too.

V. SECURITY LEVELS

The big data issues are most acutely felt in certain industries and in certain government activities. The security issues of big data systems and technologies are also applicable to cloud computing because it is very important for the network which interconnects the systems. In addition, resource allocation and memory management algorithms also have to be secure. Data mining techniques can be used in the malware detection in clouds. The challenges of security in cloud computing environments can be categorized into four levels [4][5]

(i) Network Level

(ii) User Authentication Level

(iii) Data Level

(iv) Generic issues

4.1 Network Level

The challenges that can be categorized under a network level deals with network protocols and network security, such as distributed nodes, distributed data, Internode communication.The network management problem is a very big issue in big data cloud, which is tackled using bigdata analytics.It happens because of amount of data burst in virtualized or redundant network affects the performance and security. It also makes the troubleshooting critical. Based on the changing network condition, this remains unstructured. Efficient Big Data analytic software from Apache is a Hadoop framework, includes MapReduce and the Hadoop Distributed File System (HDFS).

Using the big data applications increase the amount of real time and workload intensive transaction when massive amount of data are transferred. Because hyper scale server architectures may consist of thousands of nodes containing several processors and disks, the supporting network connecting them must be robust enough to ensure this data can move quickly and efficiently with the security. When a data load task is requested, data must be transferred and distributed to the cluster via the secured network channel. Whatever the protocols used in a cluster it should be able to handle the burst of data in normal or encrypted form without any loss or damage. At the same time by handling the security issues, the network performance and efficiency should not be affected in terms of delays in data transfer. Rightsizing switch capacity is one of the essential tasks needed to achieve network efficiency.Implementation of secure communication between nodes and applications can be done. This requires an SSL/TLS implementation that actually protects all network communications rather than just a subset. Thus the privacy of data is a huge concern in the context of Big Data.

In addition, and equally important, network performance and ease of management remain constant as the cluster scales. Because of the demands of machine-to-machine traffic flows, oversubscription should be minimized within the Big Data cluster network. The network is an essential foundation for transactions between massively parallel servers within Hadoop or other architectures and between the server cluster and existing enterprise storage systems, whether a DAS, SAN, or NAS arrangement is involved. The functionalities in the network part were discussed, hence the level of security is highly needed for Cloud Network, with the bigdata.

4.2 Authentication Level

The challenges that can be categorized under user authentication level deals with encryption/decryption techniques, authentication methods such as administrative rights for nodes, authentication of applications and nodes, and logging. The extraordinary benefits of big data are lessened by concerns over privacy and data protection. As big data expands the sources of data it can use, each data source needs to be verified and techniques should be explored in order to identify maliciously inserted data. Information security is becoming a big data analytics problem where massive amount of data will be correlated, analyzed and mined for meaningful patterns.

Security of big data can be enhanced by using the techniques of authentication, authorization, encryption and audit trails. There is always a possibility of occurrence of security violations by unintended, unauthorized access or inappropriate access by privileged users.

Some of the ways to avail authentication are

(i)Use authentication Methods such as Kerberos etc.

(ii) Encrypt the file, which provides CIA Traids of Security (Confidentiality,Integrity,Availability)

(iii)Access controls implementation: by providing privileges for user or system to enhance security.

(iv)Use Key Management, to distribute keys and certificates and manage different keys for each group, application, and user.

(v) Logging helps to detect attacks, diagnose failures, or investigate unusual behavior and activities can be recorded.

4.3 Data level

The challenges that can be categorized under data level deals with data integrity and availability such as data protection and distributed data. Many cloud environments like Hadoop store the data as it is without encryption to improve efficiency. Basically, data is processed in distributed nodes, can be anywhere across the clusters, it is very difficult to find the exact location of computation. Because of this it is very difficult to ensure the security of the place where computation is done.

Also, redundant copies of data are made to ensure data reliability. In the cloud environment, it is extremely difficult to find exactly where pieces of a file are stored. Also, these pieces of data are copied to another node/machines based on availability and maintenance operations. In traditional centralized data security system, critical data is wrapped around various security tools.Input validation and data filtering problem during data process is another issue rather then considering the security.

In order to maximize resourse utilization cloud computing helps in storing of data at a remote site. Therefore, it is very important for this data to be protected and access should be given only to authorized individuals.The Data repository should be secure then only the data will be secured. It is easy to protect data rather than the storage environment by adopting new security strategies like attribute-based encryption.

Today the data used in most big data initiatives are protected by data encryption and token-based authentication.

V. CONCULSION

Security is very important aspect in Cloud environment. The Big Data tools used to analyze the massive amount of data needs security in various levels such as Network, Authentication and Data. In this paper the three levels are studied comprehensively by focusing the works done in those levels. As a result, we conclude that the need of security various for different levels based on the process such as the data creation, data storage and transportation,data transformation and processing, and finally data usage.

REFERENCES

[1] Y, Amanatullah, Ipung H.P., Juliandri A, and Lim C. "Toward cloud computing reference architecture: Cloud service management perspective." Jakarta: 2013, pp. 1-4, 13-14 Jun. 2013.

[2] Venkata Narasimha Inukollu , Sailaja Arsi and Srinivasa Rao Ravuri,Security Issues Associated With Big Data In Cloud Computing

[3] Ren, Yulong, and Wen Tang. "A SERVICE INTEGRITY ASSURANCE FRAMEWORK FOR CLOUD COMPUTING BASED ON MAPREDUCE."Proceedings of IEEE CCIS2012. Hangzhou: 2012, pp 240 –244, Oct. 30 2012-Nov. 1 2012

[4] Hao, Chen, And Ying Qiao. "Research Of Cloud Computing Based On The HadoopPlatform.".Chengdu, China: 2011, Pp. 181 – 184, 21-23

Oct 2011.

[5] A, Katal, Wazid M, And Goudar R.H. "Big Data: Issues, Challenges, Tools And Good Practices.". Noida:2013, Pp. 404 – 409, 8-10 Aug.2013.

[6]Zhao, Yaxiong , and Jie Wu. "Dache: A data aware caching for big-data applications using the MapReduce framework." INFOCOM, 2013 Proceedings IEEE, Turin, Apr 14-19, 2013, pp. 35 - 39.

[7] Xu-bin, LI ,JIANG Wen-rui, JIANG Yi, ZOU Quan "Hadoop Applications in Bioinformatics."Open Cirrus Summit (OCS), 2012 Seventh, Beijing,Jun 19-20, 2012, pp.48 - 52.

[8] Bertino, Elisa, Silvana Castano, Elena Ferrari, and Marco Mesiti. "Specifying and enforcing access control policies for XML document sources." pp 139-151.