Challenging Security Issues Using BIG Data on Mobile Application Development

K Gurnadha Guptha[1], Vidya Sagar V Vuna[2], J.GOPAIAH [3], K Ram Mohan Goud [4], K.Anuradha[5]

[1]Assistant Professor, Computer Science Engineering,
Sri Indu College of Engg and Technology, Telangana, India
E-Mail- kbgrguptha@gmail.com

[2]Assistant Professor, Computer Science Engineering,
Sri Indu College of Engg and Technology, Telangana, India
E-Mail- vidyasagarvuna@gmail.com

[3]Assistant Professor, Computer Science Engineering,
Sri Indu College of Engg and Technology, Telangana, India
E-Mail- gopaiah1@gmail.com

[4]Assistant Professor, Computer Science Engineering,
Sri Indu College of Engg and Technology, Telangana, India
E-Mail- rammohan.koukuntla@gmail.com

[5]Assistant Professor, Computer Science Engineering,
Sri Indu College of Engg and Technology, Telangana, India
E-Mail- kanuradha.1488@gmail.com

Abstract:

The present paper highlights big data-specific security, privacy challenges, high volume inter-cloud computing, Big data, Hadoop Environment to the various rules determined by the actions taken by various organizations and companies benefit large scale and small scale industries. They promise new frontiers in science. Some companies are able to use big data for high efficiency. Security and privacy are focus by velocity, volume, and variety of big data, Cloud infrastructures, and distinct sources of data and stream of nature of acquisition. Mobile devices have virtually become many people's fifth appendage. The result is a huge flow of data from Smart-phones and tablets to organizations, which are struggling to manage and utilize that information. The traditional security mechanism are tailored to secure small-scale static opposes and stream data inadequacy. Expectations highlight challenges and bring renewed focus fortifying big data infrastructures. Cloud computing security includes rapid pace to computer security, network security, information security and data privacy. Data is protected in large and related infrastructure with policies, technologies, controls and big data tools. Mobile today can build applications which understand the users to work with the information and help in providing advice and recommendations.

Keywords:

Cloud Computing, Big Data, Hadoop, Map Reduce, HDFS (Hadoop Distributed File System)

I. Introduction:

Big data application analyzes massive amount of digital information to companies and governments collect of

the surroundings. On Daily basis there is in large 2.5 quintillion bytes of data. The data in every today is 90%. Large-scale cloud infrastructures, diversity of data sources and formats, streaming nature of data acquisition and high volume inter-cloud migration use large scale cloud infrastructures with a diversity of software platforms. These increase attacks on surface system. Traditional security mechanisms tailor secure small-scale static data inadequacy. Analytics anomaly detection generates many outliers. These are not clear to retrofit provenance and exists cloud infrastructures. Streaming data demands ultra-fast response times from security and privacy solutions.

The paper highlights top ten big data specific security and privacy challenges.

1. Secure computations in distributed programming frameworks
2. Security best practices for non-relational data stores
3. Secure data storage and transactions logs
4. End-point input validation/filtering
5. Real-time security/compliance monitoring
6. Scalable and compo-sable privacy-preserving data mining and analytics
7. Cryptographically enforced access control and secure communication
8. Granular access control
9. Granular audits
10. Data provenance

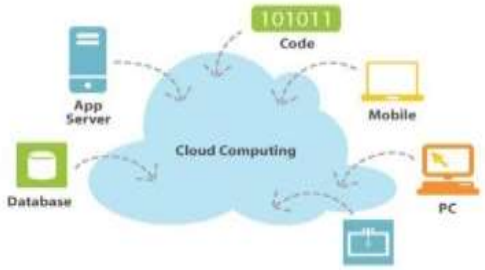
II. RELETED WORK

Cloud Computing

Cloud computing technology is basically sharing of computing resources on local /personal server to handle various applications. The word “Cloud” means “The Internet”, so Cloud Computing means a type of computing in which services are provided through Internet. The main goal of Cloud Computing is to increase the use of computing power to execute the instructions per second. Cloud Computing uses networks of large group of servers with particular connections to distribute data processing surrounded by the servers. It is replacement for installation of software suite for the computer. The Cloud Computing requires installing single software in each computer to allow the users to log into Web-based services and hosts all the programs of the user. The user's end run’s the cloud interface software to connect the cloud. Cloud Computing is contained in a front end and back end.

The front end includes the user’s computer and software to access the cloud network. Back end consists of various computers, servers and database systems that create the cloud. User’s can access applications in the

cloud network from anywhere connecting the cloud using Internet. Some of the real time applications used is Cloud Computing are Gmail, Google Calendar, Google Docs and Drop box etc.,



Big Data

Big Data is the word used to describe 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 technologies. The term "Big Data " is believed to be originated from the Web search companies who had to query loosely structured very large distributed data. The three e main terms that signify Big Data have the following properties:

a) Volume: Many factors contribute towards increasing Volume - storing transaction data, live streaming data and data collected from sensors etc.,
b) Variety: Today data comes in all types of formats – from traditional databases, text documents, emails, video, audio, transaction s etc.,

c) Velocity: This means how fast the data is being produced and how fast the data needs to be processed to meet the demand. The other two dimensions that need to consider with respect to Big Data are Variability and Complexity.

d) Variability: Along with the Velocity, the data flows can be highly inconsistent with periodic peaks.

e) Complexity: Complexity of the data also needs to be considered when the data is coming from multiple sources. The data must be linked, matched, cleansed and transformed into required formats before actual processing.

Technologies today not only support the collection of large amounts of data, but also help in utilizing such data effectively. Some of the real time examples of Big Data are Credit card transactions made all over the world with respect to a Bank, Walmart customer transactions, and Facebook users generating social interaction data.
When making an attempt to understand the concept of Big Data, the words such as "Map Reduce" and "Hadoop" cannot be avoided

BIG DATA USES VARIOUS FILEDS

Sample Big Data Use Cases Industry Government Banking and Finance Information Technology

Telecommunications Health Care and Life Sciences Sample Use Cases Fraud detection Compliance and regulatory analysis Climate analysis/weather prediction Compliance and regulatory analysis Risk analysis and management Fraud detection and security analytics CRM and customer loyalty programs Credit risk scoring and analysis High-speed arbitrage trading Trade surveillance Abnormal trading pattern analysis Information security analytics Threat intelligence Log management Revenue assurance and price optimization Customer churn prevention Campaign management and customer loyalty Call detail record (CDR) analysis Network performance and optimization Mobile user location analysis Clinical trials data analysis Disease pattern analysis Campaign and sales program optimization Medicaid and Medicare fraud Patient care quality and program analysis Population health Medical device and pharmacy supply chain management Drug discovery and development analysis

Hadoop

Hadoop, which is a free, Java-based programming framework, supports the processing of large sets of data in a distributed computing environment. It is a part of theApache project sponsored by the Apache Software Foundation. Hadoop cluster uses a Master/Slave structure. Using Hadoop, large data sets can be processed across a cluster of servers and applications can be run on systems with thousands of nodes involving thousands of terabytes. Distributed file system in Hadoop helps in rapid data transfer rates and allows the system to continue its normal operation even in the case of some node failures. This approach lowers the risk of an entire system failure, even in the case of a significant number of node failures. Hadoop enables a computing solution that is scalable, cost effective, and 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 projects – Map Reduce and Hadoop Distributed File System (HDFS).

Map Reduce

Hadoop Map Reduce is a foundation which prepares applications to process huge volume of data in parallel clusters. The commodity hardware resources are reliable in fault-tolerant manner. The Map Reduce job divides the data into individual chunks which are processed by Map jobs in parallel. The outputs of the maps are sorted by foundation with input to reduce tasks. The input and output jobs are stored in a file-system. Scheduled, Monitor and launch are failed tasks which are taken care in the framework.

Hadoop Distributed File System (HDFS)

HDFS file system which spans the nodes in Hadoop cluster for data storage. It advances together file systems on local nodes to make them a large file system. HDFS promotes reliability by replicating data across multiple sources to overcome Galera Documentation.

Big data applications

The big data application refers to the large scale distributed applications which usually work with large data sets. Data exploration and analysis turned into a difficult problem in many sectors in the span of big data. With large and complex data, computation becomes difficult to be handled by the traditional data processing applications which triggers the development of big data applications .Google's map reduce framework and apache Hadoop are the defector software systems for big data applications, in which these applications generates a huge amount of intermediate data. Manufacturing and Bioinformatics are the two major areas of big data applications.

Big data provides infrastructure for transparency in production and manufacturing which is able to unravel uncertainties like inconsistent component performance and availability. Big Data applications which are conceptual framework to predictive manufacture initiates the data acquisition as a possibility to acquire

different types of sensory data such as pressure, vibration, acoustics, voltage, current and control data. Big data manufacturing constructs Sensory data and historical data. Big data generates combination acts of input to predictive tools and prevent strategies say prognostics and health management.

Hadoop Application in Bioinformatics covers next generation sequence and other biological domains. Bioinformatics requires large scale data to analyze - Hadoop. Cloud computing gets parallel distributed computing framework together with computer clusters and web interfaces.

App development tool helps news papers stay with the times.

Online Media overtakes pulse of community and gathers reporting to the news for internal operations. Gmail, Google Docs and bunch of Google apps are assumptions down the Google path and turn to right choice.

Big data advantages

Big data software package provides rich set of tools and options to individuals who map the complete data landscape across the company allowing individuals to analyze threats to face internally. These are advantages to big data and are of safe data. One can detect the potential sensitive information which is not protected in appropriate manner and stores regulatory requirements.

Common characteristics of big data are

a) Big data assembles structured and unstructured data.
b) Addresses speed and scalability, mobility and security, flexibility and stability.
c) Big data realization time informs the critical extraction value of different data sources, added mobile devices, radio frequency identification, web and list of automated sensory technologies.

Organizations and business benefit speed, capacity and scalability of cloud storage. End users visualize the data and companies to find new business opportunities. There is notable advantage with big-data, data analytics, and individuals to personalize content or look and feel of website in real time to suit the customer with website. Big data combines predictive analytics, produces challenges many industries. The results explore in four areas:

a) Study risks of large portfolios.
b) Detect, prevent and re-audit financial fraud
c) Improve delinquent collections
d) Execute high value market campaigns

Security in Big data

Many businesses use big data for marketing and research but they do not have fundamental assets particularly

for security perspective. A security breach occurs to big data and results in more serious legal repercussions and reputation damages that are present at times. Many companies use technology to store and analyze bytes of data about the company, business and customers. Information classified for even more critical and makes big data secure, techniques with encryption, logging, honey pot detection. Many organizations deploy big data for fraud detection.

Advanced threats challenge detection and prevent the malicious intruders to solve big data style analysis. These techniques detect the threats of early stages and use more sophisticated pattern analysis of multiple data. Security and data privacy challenges the existence of industries and federal organizations. The increase in use of big data in business, many companies wrest with privacy. Data privacy is liable for the companies on privacy defense. Security and privacy consider the asset; a selling point exists for customers and stakeholders with a balance between data privacy and national security.

Development of Mobile Application by Big Data:

1. App development tool serve news papers to stay at times.
2. No additional features required for mobile design and advancement.
3. Active use of Mobile BYOD.
4. Integrated Software's and infrastructures give next generation Apps.
5. Codeless covers IT in abundant ground.
6. Security and monitor capabilities discover Enterprise Applications.
7. Risks can be managed to public and private API's.

ISSUES AND CHALLENGES

Cloud computing through numerous security issues and encompasses more technologies to include networks, databases, operating systems, virtualization, resource schedule, transaction management, load balancing, concurrency control and memory management.
The security issues of systems and technologies apply to cloud computing.
Network interconnectivity systems in cloud based software is highly important for the application Virtual paradigm is cloud computing and results in several security concerns and are map to virtual machines and physical machines to perform the security.

The challenges of security in cloud computing environments categorize into network level, user authentication level, data level, and generic issues.

Network level: The challenges are categorized under network level deal with network protocols and network security, like distributed nodes, distributed data, Internodes communication.

Authentication level: Challenges categorized authenticate level deals with encryption/decryption techniques, authentication methods with administrative rights for nodes, applications and nodes, and logs.

Data level: The challenges are categorized under data level deals with data integrity and availability like data protection and distributed data.

Generic types: The challenges are categorized under general level with traditional security tools applying Different technologies

III. PROPOSAL SOLUTION:

Various security measures of security of cloud computing environment are build with the mixture of many technologies, proposing various solutions collectively to make environment secure. The proposed solutions encourage multiple technologies and tools to mitigate the security problem that was specified. The recommendations of Security design stabilize the efficiency and scale of cloud systems.

Security measures ensure security in cloud environment.

File Encryption

Data is present in machines in cluster and a hacker steals the critical information. All the data stored are encrypted with different encryption keys to use different machines and key information is stored centrally with strong firewalls. A hacker is able to get data and lessens extraction of meaningful information from misuse. User data is secure and encrypted.

Network Encryption

The network communication encrypts industry standards. RPC procedure takes place over SSL for the even to hack to tap the network communication packets that cannot extract to useful information or manipulate the packets.

Logging

Map reduce jobs modify data to log. Information for users is responsible for the jobs and is logged with many logs. Any malicious operations performed are manipulated the data in nodes.

Software Format and Node Maintenance

Nodes that run software are formatted regularly to eliminate the virus present. All the application software's and Hadoop software are updated to make system more secure.

Nodes Authentication

Nodes that join the cluster should authenticate the malicious node. The nodes are allowed to join the cluster. Authentication techniques like Kerberos use to validate the authorized nodes from malicious nodes.

Rigorous System Testing of Map Reduce Jobs

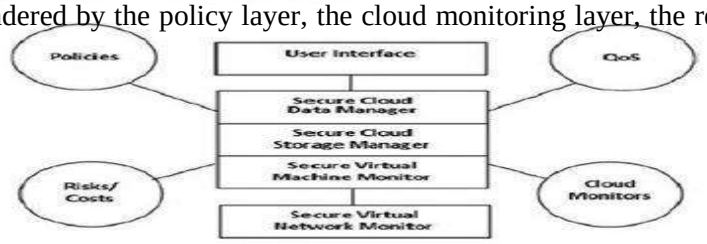
Once Developer writes a map to reduce a job they are thoroughly tested in distributed environment instead of a single machine to ensure robustness and stability of job.

Honeypots Nodes

Honey pot nodes are present in cluster to appear in a regular node but are a trap. These honeypots trap the hackers and necessary actions would be taken to eliminate hackers.

Layered Framework for Assuring Cloud

A layered framework for assuring cloud computing as shown in Figure consists of the secure virtual machine layer, secure cloud storage layer, secure cloud data layer, and the secure virtual network monitor layer. Cross cutting services are rendered by the policy layer, the cloud monitoring layer, the reliability layer and the risk analysis layer.



\

Fig: Layered framework for assuring cloud

Third Party Secure Data Publication to Cloud

Cloud computing helps in storing of data at a remote site in order to maximize resource utilization. It is very important for data to be protected and access should be given only to authorized individuals. therefore this fundamentally amounts to secure third party publication of data that is required for data outsourcing, as well as for external publications. In the cloud environment, the machine serves the role of a third party publisher, which stores the sensitive data in the cloud. This data needs to be protected, and the above discussed techniques have to be applied to ensure the maintenance of authenticity and completeness.

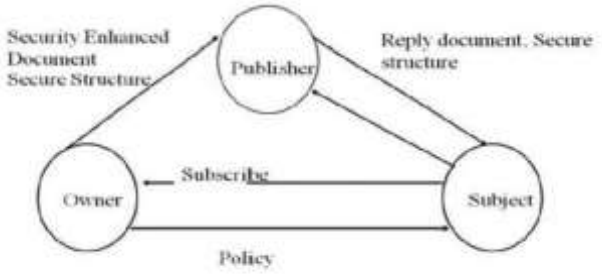


Fig: Third party secure data publication applied to cloud.

Access Control

Integration of mandatory access controls and differential privacy in distributed environment will be a high-quality security measure. Data providers will control the security policy of their sensitive data. Also control the mathematical bound on privacy violation that could take place. In the above approach, users can perform data computation without any leakage of data. To prevent information leak, SELinux will be used. SELinux is nothing but Security-Enhanced Linux,

Enforcement of differential privacy will be done using modification to Java Virtual Machine and the Map Reduce framework. They have inbuilt applications which store the user identity pool for the whole cloud service. As a result the cloud service will not have to maintain each user's identity for each application. In addition to the above methodologies, cloud service will support third party authentication. The third party will be trusted by both the cloud service and accessing user. Third party authentication will add an additional security layer to the cloud service.

Real time access control will be a good security measure in the cloud environment. In adding together to access control to the cloud environment, operational control within a database in the cloud can be used to

prevent configuration drift and unauthorized application changes. Multiple factors such as IP address, time of the day, and authentication method can be used in a flexible way to employ above measures. For example, access can be restricted to specific middle tier, creating a trusted path to the data. Keeping a security administrator separate from the database administrator will be a good idea. The label security method will be implemented to protect sensitive data by assigning data label or classifying data.

Data preserve be classified as public, confidential and sensitive. If the user label matches with the label of the data, then access is provided to the user. Examination of numerous data breaches has shown that auditing could have helped in early detection of problems and avoids them. Auditing of events and tracking of logs taking place in the cloud environment will enable possible attack. Fine grain auditing just like Oracle 9i enables conditional auditing on the specific application column.

IV. CONCLUSION

Mobile Application Development in Cloud environment is widely used in industry and research aspects; therefore security is an important aspect for organizations running on these cloud environments. Using proposed approaches, cloud environments can be secured for complex business operations.

REFERENCES

1. 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

2. 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.

3. 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.

4. Lu, Huang, Ting-tin Hu, and Hai-shan Chen. "Research on Hadoop Cloud Computing Model and its Applications.". Hangzhou, China: 2012, pp. 59 – 63, 21-24 Oct. 2012.

5. K, Chitharanjan, and Kala Karun A. "A review on hadoop — HDFS infrastructure extensions.". JeJu Island: 2013, pp. 132-137, 11-12 Apr. 2013.

6. F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C. "Business model canvas perspective on big data applications." *Big Data, 2013 IEEE International Conference*, Silicon Valley, CA, Oct 6-9, 2013, pp. 32 - 37.
7. 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.

8. P.R , Anisha, Kishor Kumar Reddy C, Srinivasulu Reddy K, and Surender Reddy S. "Third Party Data Protection Applied To Cloud and Xacml Implementation in the Hadoop Environment With Sparql."2012. 39-46, Jul – Aug. 2012.

BIOGRAPHIES

K Gurnadha Guptha, working as Assistant Professor in Sri Indu College of Engg and Technology He has done a post-graduate from Swarnandhra College of Engineering and Technology, Narasapur, West Godavari,AP. He has done a graduate from ANNA UNIVERSITY, CHENNAI,TN. His main research interests are Data warehousing and Mining, Distributed Database System.

Vidya sagar V Vuna , working as Assistant Professor in Sri Indu College of Engg and Technology He has done a post-graduate from GITAM UNIVERSITY ,VISAKAPATNAM., AP.He has total teaching experience of 5+ years His main research interests are Data warehousing and Mining, Distributed Database System.

J .Gopaiah B.Tech Computer Science Engineering From Lords Institute Of Engg And Technology . M.Tech Computer Science From Nishitha Institute Of Engg And Technology. Currently He Working As Asst Prof In Sri Indu College Of Engineering & Technology Having One Year Of Academic Experience. His Interested Subjects Include Data Mining Software Engineering Network Security, Big Data & Cloud Computing.

K RAM MOHAN GOUD , working as Assistant Professor in Sri Indu College of Engg and Technology He has done a post-graduate from He has done a graduate from TKR College of Engineering and Technology Hyderabad, Telangana. has done a graduate from **Anurag Engineering College Nalgonda,** Telangana. He has total teaching experience of 5+ years His main research interests are Data warehousing and Mining, Distributed Database System.

K.Anuradha, working as Assistant Professor in Sri Indu College of Engg and Technology She has done a post-graduate from She has done a graduate from Bhoj Reddy Engg college for women, Telangana. has done a graduate Malla Reddy Engg college of technology Telangana. . She has total teaching experience of 5+ years His main research interests are Data warehousing and Mining, Distributed Database System.