© 2016 IJSRSET | Volume 2 | Issue 1 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Big Data Security Issues in Networking

A. M. Chandrasekhar[1], Jagadish Revapgol[2], Vinayaka Pattanashetti[3]
[1]Assistant Professor, [2 & 3]M.Tech Students,
Department of Computer Science & Engineering, SJCE,
Mysuru, Karnataka, India

ABSTRACT

As of now with the advantages of big data in many fields like Business, E-Commerce, Social Media, Networking and so on, approach in this paper concentrates on security issues which our future is going to face if they are not encountered today, especially in security of public, private data. Data may be available publically to all or it may be some confidential known to very few systems or persons. Big data technology makes use of massive datasets being flown through social media websites and many other sources, analyses it and make pro-intelligent decisions i.e., immature output that is not completely accepted it may violate the privacy concern of a company or system or it may be an individual. So it's our today's responsibility to maintain data confidentiality and data integrity together so that we will not face such problems in future. We propose some of important, major security issues that will emerge today or tomorrow.

Keywords : Hadoop, Map Reduce, Network Encryption and Zettaset Orchestrator.

I. INTRODUCTION

As keeping security of network in mind we are going to introduce something regarding usage of Big Data in networking today as well as tomorrow. Here are some buzz words of Big Data that themselves make sense the existence of security holes if we are supposed to apply Big Data Analytics in Network Security.

A. Data Leakage

One of the major security hole of big data technology is data leakage. Availability of massive data, increased rate of sharing of data, globalization of information, and mainly absence of security policies and procedures makes it difficult to have control over flow of data across galaxy of internet. Information leaked is irrespective of confidentiality that data may be highly sensitive, very confidential, can be public. [1].

B. Undefined Source

Source of big data, as we know, are web (text data from open social web platforms like Facebook, Orkut, Twitter, etc.), video and audio data and image files. For a particular big data application data that is input to application cannot be constrained on the basis of privacy concern [6-8]. Although we know the abstract source of data, we cannot say, for example this particular part of data stream is generated by a particular person's chat history or his cookie or web log from this particular site. Some of example of sources are given below in Table I Although we know that exact source of data from where we are going to access data may not satisfy the privacy policy of that particular source. If measures are not taken place this kind of activity may be lead to an offensive.

The Big Data landscape is incredibly diverse across three areas: Data form, Data Sources and Data Consumers as shown in Table I

The Big Data landscape is incredibly diverse across three areas: Data form, Data Sources and Data Consumers as shown in Fig 1

II. 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 the Apache project sponsored by the Apache Software Foundation [9-11]. 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 [12-13]. 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) [2].

TABLE I
DIVERSION OF BIG DATA ACROSS THREE MAJOR AREAS

Data Form
May be structures, like databases and transactional data, or it could be unstructured. Think Office documents, images, and raw data stored as flat files
Data Sources
Include financial accounting applications, sales and product data, CRM Applications, email files, server logs, office files, images, mobile device data including geo-location and much more
Data Consumers
Range from department level analysts to senior business managers to IT and Information- Security teams to partners, customers and various business users

A. Map Reduce

Hadoop Map Reduce is a framework used to write applications that process large amounts of data in parallel on clusters of commodity hardware resources in a reliable, fault-tolerant manner. A Map Reduce job 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. [14-15]

B. Hadoop Distributed File System (HDFS)

HDFS is a file system that spans all the nodes in a Hadoop cluster for data storage. 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.

For marketing and research, many of the businesses uses big data, but may not have the fundamental assets particularly from a security perspective. If a security breach occurs to big data, it would result in even more serious legal repercussions and reputational damage than at present[16]. In this new era, many companies are using the technology to store and analyze petabytes of data about their company, business and their customers. As a result, information classification becomes even more critical. For making big data secure, techniques such as encryption, logging, and honey-pot detection must be necessary. In many organizations, the deployment of big data for fraud detection is very attractive and useful.[3]

The challenge of detecting and preventing advanced threats and malicious intruders must be solved using big data style analysis. These techniques help in detecting the threats in the early stages using more sophisticated pattern analysis and analyzing multiple data sources[17].

Not only security but also data privacy challenges existing industries and federal organizations. With the increase in the use of big data in business, many companies are wrestling with privacy issues. Data privacy is a liability, thus companies must be on privacy defensive[18]. But unlike security, privacy should be considered as an asset, therefore it becomes a selling point for both customers and other stakeholders. There should be a balance between data privacy and national security.

C. File Encryption

Since the data is present in the machines in a cluster, a hacker can steal all the critical information. Therefore, all the data stored should be encrypted[19]. Different encryption keys should be used on different machines and the key information should be stored centrally behind strong firewalls. This way, even if a hacker is able to get the data, he cannot extract meaningful information from it and misuse it. User data will be stored securely in an encrypted manner.

D. Network Encryption

All the network communication should be encrypted as per industry standards. The RPC procedure calls which take place should happen over SSL so that even if a hacker can tap into network communication packets, he cannot extract useful information or manipulate packets[20-21].

E. Logging

All the map reduce jobs which modify the data should be logged. Also, the information of users, which are responsible for those jobs should be logged. These logs should be audited regularly to find if any, malicious operations are performed or any malicious user is manipulating the data in the nodes.[4]

F. Software Format and Node Maintenance

Nodes which run the software should be formatted regularly to eliminate any virus present. All the application softwares and Hadoop software should be updated to make the system more secure[22-24].

III. NODES AUTHENTICATION

Whenever a node joins a cluster, it should be authenticated. In case of a malicious node, it should not be allowed to join the cluster. Authentication techniques like Kerberos can be used to validate the authorized nodes from malicious ones[25-27].

A. Rigorous System Testing of Map Reduce Jobs

After a developer writes a map reduce job, it should be thoroughly tested in a distributed environment instead of a single machine to ensure the robustness and stability of the job.It can be executed on a test cluster to identify potential integration and scaling issues. Or, the Hadoop classes MiniDFSCluster and MiniMRCluster could be leveraged tocreate additional tests that execute against a pseudo-cluster[28].

B. Solution: Move Security Closer to the Data

A Forrester report, the “Future of Data Security and Privacy: Controlling Big Data”, observes that security professionals apply most controls at the very edges of the network. However, if attackers penetrate your perimeter, they will have full and unrestricted access to your big data. The report recommends placing controls as close as possible to the data store and the data itself, in order to create a more effective line of defence. Thus, if the priority is data security, then the cluster must be highly secured against attacks[29].

C. Deploy a Purpose-Built Security Solution for Hadoop and Big Data

Only a new approach that addresses the unique architecture of distributed computing can meet the security requirements of the enterprise data center and the Hadoopcluster environment[30].

“Only a new approach that addresses the unique architecture of distributed computing can meet the security requirements of the enterprise data center and the Hadoop cluster environment.” Zettaset Orchestrator provides an enterprise-class security solution for big data that is embedded in the data cluster itself, moving security as close to the data as possible, and providing protection that perimeter security devices such as firewalls cannot deliver[31].

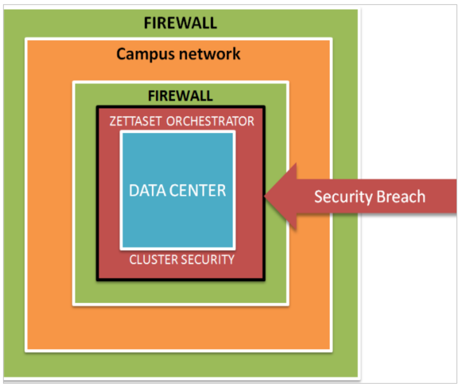

Figure 1: Zettaset Orchestrator provides security from within the data center cluster. Even if perimeter security is breached, the cluster and sensitive data are still protected by Orchestrator’s comprehensive security wrapper.

At the same time, Orchestrator addresses the security gaps that open-source solutions typically ignore, with a comprehensive big data management solution which is hardened to address policy, compliance, access control and risk management within the Hadoop cluster environment.[5]Orchestrator includes RBAC, which significantly strengthens the user authentication process. Orchestrator simplifies the integration of Hadoop clusters into an existing security policy framework, with support for LDAP and AD. For those organizations with compliance reporting requirements, Orchestrator includes extensive logging, search, and auditing capabilities[32].

Orchestrator addresses the critical security gaps that exist in today's distributed big data environment with these capabilities:

- Fine-grained Access Control – Orchestrator significantly improves the user authentication process with RBAC.
- Policy Management – Orchestrator simplifies the integration of Hadoop clusters into an existing security policy framework with support for LDAP and AD[33].
- Compliance Support – Orchestrator enables Hadoop clusters to meet compliance requirements for reporting and forensics by providing centralized configuration management, logging, and auditing. This also enhances security by maintaining tight control of ingress and egress points in the cluster and history of access to data.

Zettaset Orchestrator is the only solution that has been specifically designed to meet the security requirements of the distributed architectures which predominate in big data and Hadoop environments. Orchestrator creates a security wrapper around any Hadoop distribution and distributed computing environment, making it enterprise-ready[34].

With Orchestrator, organizations can now confidently deploy Hadoop in data the center environments where security and compliance is a business imperative."Zettaset Orchestrator is only solution that has been specifically designed"

IV. CONCLUSION

By this paper work we would like to conclude that security of network that makes use of Big Data technology must be more secure in order to enhance our vision in network security that will be used in integration with Big Data [35]. So, as security is very basic and fundamental need we must be aware of security violations in future.

V. REFERENCES

[1] N, Gonzalez, Miers C, Redigolo F, Carvalho T, Simplicio M, de Sousa G.T, and Pourzandi M. "A Quantitative Analysis of Current Security Concerns and Solutions for Cloud Computing.".

[2] Bamford, J. (2013). Five myths about the National Security Agency. The Washington Post. http://articles.washingtonpost.com/2013-06-21/opinions/40114085_1_national-security-agency-foreign-intelligence-surveillance-court-guardian.

[3] Bamford, J. (2012). The NSA is building the country's biggest spy center (watch what you say). WIRED. http://www.wired.com/threatlevel/2012/03/ff_nsadatacenter.

[4] Barker, M., & Reed, M. S. C. (2013). A research environment for high risk data. Presented at the Research Data Management Implementations Workshop. Chicago, IL, USA: The University of Chicago. http://rdmi.uchi-cago.edu/sites/rdmi.uchicago.edu/files/uploads/Barker,MandReed,M_AResearchEnvironmentforHighRiskData.pdf.

[5] DBMS2. (2009a). Followup on IBM System S/InfoSphereStreams.DBMShttp://www.dbms2.com/2009/05/18/followup-on-ibm-system-infosphere-streams

[6] A M. Chandrashekhar, K. Raghuveer, "Fusion of Multiple Data Mining Techniques for Effective Network Intrusion Detection – A Contemporary Approach", Proceedings of the 5th International Conference on Security of Information and Networks (SIN 2012), 2012, pp 33-37.

[7] A M. Chandrashekhar, K. Raghuveer, "An Effective Technique for Intrusion Detection using Neuro-Fuzzy and Radial SVM Classifier", The Fourth International Conference on Networks & Communications (NetCom-2012), 22~24, Dec-2012.

[8] A M. Chandrashekhar, K. Raghuveer , "Intrusion Detection Technique by using K-means, Fuzzy Neural Network and SVM classifiers", 2013 IEEE International Conference on Computer Communication and Informatics (ICCCI -2013), 4~06, Jan2013,

[9] A M. Chandrashekhar, K. Raghuveer, "Confederation of FCM Clustering, ANN and SVM Techniques of Data mining to Implement Hybrid NIDS Using Corrected KDD Cup Dataset", IEEE International Conference on Communication and Signal Processing (ICCSP),2014, pp 672-676.

[10] A M Chandrashekhar, K. Raghuveer, "Hard Clustering Vs. Soft Clustering: A Close Contest for Attaining Supremacy in Hybrid NIDS Development", Proceedings of International

Conference on Communication and Computing (ICC - 2014), Elsevier science and Technology Publications.

[11] A M. Chandrashekhar, K. Raghuveer, "Amalgamation of K-means clustering algorithem with standard MLP and SVM based neural networks to implement network intrusion detection system", Advanced Computing, Networking, and Informatics –Volume 2(June 2014), Volume 28 of the series Smart Inovation, Systems and Technologies pp 273-283.

[12] A M Chandrashekhar, K. Raghuveer, "Diverse and Conglomerate Modi-operandi for Anomaly Intrusion Detection Systems", International Journal of Computer Application (IJCA) Special Issue on "Network Security and Cryptography (NSC)", 2011.

[13] A M. Chandrashekhar, K. Raghuveer, "Performance evaluation of data clustering techniques using KDD Cup-99 Intrusion detection data set", International Journal of Information and Network Security (IJINS), ISSN: 2089-3299, Vol-1, No.4, October 2012, pp. 294~305.

[14] A M. Chandrashekhar, K. Raghuveer, "Fortification of hybrid intrusion detection system using variants of neural networks and support vector machines", International Journal of Network Security & Its Applications (IJNSA) ISSN: 0974-9330[online] & 0975-2307[print].Vol.5, Number 1, January 2013.

[15] A. M. Chandrashekhar, K. Raghuveer , "Improvising Intrusion detection precision of ANN based NIDS by incorporating various data Normalization Technique – A Performance Appraisal", International Journal of Research in Engineering & Advanced Technology(IJREAT), Volume 2, Issue 2, Apr-May, 2014.

[16] A. M Chandrashekhar, Puneeth L Sankadal, Prashanth Chillabatte, "Network Security situation awareness system", International Journal of Advanced Research in Information and Communication Engineering(IJARICE), Volume 3, Issue 5, May 2015.

[17] A.M.Chandrashekhar, Prashanth G M, Anjaneya Bulla, "Secured infrastructure for multiple group communication" International Journal of Advanced Research in Information and Communication Engineering (IJARICE), Volume 3, Issue 5, May 2015.

[18] A. M. Chandrashekhar, Sowmyashree K.K, Sheethal R.S, "Pyramidal aggregation on Communication security" International Journal of Advanced Research in Computer Science and Applications (IJARCSA), Volume 3, Issue 5, May 2015.

[19] A .M. Chandrashekhar, Huda Mirza saifuddin, Spoorthi B.S, "Exploration of the ingredients of original security" International Journal of Advanced Research in Computer Science and Applications(IJARCSA), Volume 3, Issue 5, May 2015.

[20] A. M.Chandrashekhar, Syed Tahseen Ahmead, Rahul N, "Analysis of Security Threats to Database Storage Systems" International Journal of Advanced Research in data mining and Cloud computing (IJARDC), Volume 3, Issue 5, May 2015.

[21] A.M.Chandrashekhar, Sachin Kumar H S, Yadunandan, "Advances in Information security risk practices" International Journal of Advanced Research in data mining and Cloud computing (IJARDC), Volume 3, Issue 5 May 2015.

[22] A. M. Chandrashekhar, Madhura S Hegde, Aarabhi Putty, "A Survey: Combined impact of cryptography and steganography" International Journal of Engineering Research (IJOER), Volume 3, Issue 5, May 2015.

[23] A.M.Chandrashekhar, Koushik P, JagadeeshTakkalakaki, "Information security threats, awareness and coginizance" International Journal for Technicle research in Engineering (IJTRE), Volume 2, Issue 9, May 2015.

[24] A. M. Chandrashekhar, Rahil kumar Gupta, Shivaraj H. P, "Role of information security awareness in success of an organization" International Journal of Research(IJR) Volume 2, Issue 6, May 2015

[25] A. M. Chandrashekhar, Arpitha, Nidhishree G, "Efficient data accessibility in cloud with privacy and authenticity using key aggregation cryptosystem", International Journal for Technological research in Engineering (IJTRE), Volume 3, Issue 5, JAN-2016.

[26] A. M. Chandrashekhar, Hariprasad M, Manjunath A, "The Importance of Big Data Analytics in the Field of Cyber Security", International journal of scientific Research and Development (IJSRD),Volume 3, Issue 11, JAN-2016.

[27] A. M. Chandrashekhar, Chitra K V, Sandhya Koti, "Security Fundamentals of Internet of Things",*International Journal of Research (IJR), Volume 3, Issue no1, JAN-2016.*

[28] .A. M. Chandrashekhar, Anjana D, Muktha G, "Cyber stalking and Cyber bullying: Effects and prevention measures", Imperial Journal of Interdisciplinary Research (IJIR), Volume 2, Issue 2, JAN-2016.

[29] A. M. Chandrashekhar, Sahana K, Yashaswini K, "Securing Cloud Environment using Firewall and VPN", "International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Volume 6, Issue-1, January-2016

[30] A. M. Chandrashekhar, Sahana K, Yashaswini K, "Securing Cloud Environment using Firewall and VPN", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Volume 6, Issue-1, January-2016.

[31] A. M Chandrashekhar,Puneeth L Sankadal , Prashanth Chillabatte, "Network Security situation awareness system" International Journal of Advanced Research in Information and Communication Engineering(IJARICE), Volume 3, Issue 5, May 2015.

[32] A. M. Chandrasekhar, Jagadish Revapgol, Vinayaka Pattanashetti, "Security Issues of Big Data in Networking", International Journal of Scientific Research in Science, Engineering and Technology (*IJSRSET),* Volume 2, Issue 1, JAN-FEB,2016.

[33] A. M Chandrashekhar, Lavanya C P, Ramya J, "Detection of Phishing Websites", International journal of Advanced research in information and communication(IJARIC),Volume 4, Issue 1,Jan-2016

[34] A. M. Chandrasekhar, NgaveniBhavi, Pushpanjali M K, "Hierarchical Group Communication Security", International journal of Advanced research in Computer science and Applications(IJARCSA), Volume 4, Issue 1,Jan-2016

[35] A. M. Chandrasekhar, Vasudeva, Danish Pasha, Securing Cloud using Public Key Infrastructure, International journal of Advanced research in Data mining and cloud computing (IJARDC), Volume 4, Issue 1, Jan-2016.