Asian Journal of Applied Science and Technology (AJAST)
Volume 1, Issue 5, Pages 113-116, June 2017

A Survey on Big Data Security and Related Techniques to Improve Security

S.Castro[1] and R.Pushpalakshmi[2]

[1]*Department of Information Technology, Kalasalingam Institute of Technology, Tamilnadu, India. Email: suseelcastro@gmail.com*
[2]*Department of Information Technology, PSNA College of Engineering and Technology, Tamilnadu, India. Email: pushparaman@rediffmail.com*

Article Received: 18 May 2017 *Article Accepted: 04 June 2017* *Article Published: 07 June 2017*

ABSTRACT

Enhancement of security and privacy in mobile data centers is challengeable with efficient security key management. For marketing and research, many of the businesses use big data, but may not have the fundamental assets particularly from a security perspective. Security of Big Data is a big concern. In broad sense Big Data contains two types of data such as structured and unstructured. To provide security to unstructured data is more difficult than that of structured. In this paper deals with a survey on big data security and proposed new technique called big data security suite with quantum cryptographic key management scheme. The proper security standards of algorithms can be activated using and which has been interfaced with the data node. Unstructured data by considering the types of the data and their sensitivity levels. We have reviewed the different analytics methods of Big Data, which gives us the facility to build a data node of databases of different types of data. Each type of data has been further classified to provide adequate security and enhance the overhead of the security system.

Keywords: Security, Authentication, Information secrecy, Data node, NoSql, Hadoop and MapReduce.

1. INTRODUCTION

On the off chance that a security break jumps out at enormous information, it would result in considerably more genuine legitimate repercussions and reputational harm than at present. In this new time, many organizations are utilizing the innovation to store and examine petabytes of information about their organization, business and their clients. Therefore, data grouping turns out to be significantly more basic. For making enormous information secure, strategies, for example, encryption, logging, honeypot identification must be essential. In numerous associations, the arrangement of huge information for misrepresentation location is exceptionally appealing and helpful. The test of distinguishing and forestalling propelled dangers and malignant gate crashers, must be understood utilizing enormous information style examination. These procedures help in recognizing the dangers in the early stages utilizing more advanced example examination and breaking down various information sources [1]. Security as well as information protection challenges existing enterprises and government associations. With the expansion in the utilization of huge information in business, many organizations are grappling with protection issues. Information security is a risk, in this manner organizations must be on protection cautious. Be that as it may, not at all like security, protection ought to be considered as an advantage, in this way it turns into an offering point for both clients and different partners. There ought to be a harmony between information protection and national security. In Big Data with its volume there are assortments of information, for example, organized, semi-organized, unstructured and so forth. The security of organized information can be considered by existing security philosophies or gauges utilizing SQL questions. Along these lines, our advantage is to give security to unstructured information, which incorporate content, XML, picture, video, sound et cetera. We have built up an approach for security of unstructured Big Data by considering the current measures or calculations of security. This approach portrays examination of the unstructured information utilizing information investigation innovations; manufacture an information hub of databases. In this paper concentrate on various overviews in huge information security issues.

2. SURVEY ON BIG DATA SECURITY

2.1 Big data causes privacy violation in various applications

On account of the health care enormous points of interest in securing the soundness of patients, huge information is exceedingly bolstered by medicinal services framework. Enormous information data is utilized to perceive individuals with a high danger of certain therapeutic conditions at early stage and giving enhanced quality care and bringing down the expansion cost of human services. Despite the fact that there are colossal advantages, new reviews are uncovering that huge information might be less secure than at first idea. According to overview it is discovered that, however the social insurance information is close to home, it is effortlessly open. It is critical to be cognizant about security and protection suggestions taking advantage of enormous information. Huge information permits the expectation of a considerable amount of other data about individuals. The data enormous information can anticipate is progressively building up the possibility to be utilized as a method for oppressing individuals in [7] an assortment of socioeconomics. A review demonstrates that when perception of status i.e. like data from face book was dissected, it gave precise data to separate men contingent upon race, liquor utilization, sexual orientation and so forth. It is particularly worried by many individuals that associations, managers, training framework may utilize such models and begin segregating individuals in light of numerous human arranged parameters. One of the real uses of huge information is promoting where the advertisers attempt to put their items and administrations before profoundly focused on clients. In any case, when the client is classified into one classification in view of their practices, there is probability for damage. Disregarding the likelihood for damage, advertisers still utilize huge information to go for individuals via web-based networking media stages like web indexes and email. Mighty passage into individual range by

giving promotions in view of companions, likes and email substance is bringing on nervousness among shoppers.

2.2 Big data security challenges in cloud

Secure calculations in appropriated programming structures. The primary recognized hazard delves into the security of computational components in structures, for example, MapReduce, with two particular security concerns delineated. To start with, the dependability of the "mappers," which are the code that breaks information into pieces, dissects it and yields key-esteem sets, should be assessed. Second, information cleansing and de-recognizable proof capacities should be actualized to keep the capacity or spillage of delicate information from the stage ought to be executed through information sharpening what's more, de-recognizable proof. Undertakings utilizing complex apparatuses, for example, MapReduce should utilize devices, for example, Mandatory Get to Controls inside SELinux and de-identifier schedules to finish this; on a similar note, undertakings ought to ask with respect to how cloud suppliers are controlling and remediating this issue in their surroundings. Security best practices for non-social information stores. The utilization of No SQL and other expansive scale, non-social information stores may make new security issues because of a conceivable absence of capacities in a few crucial territories, including any genuine confirmation, encryption for information very still or in travel, logging or information labeling, and grouping. Associations need to consider the utilization of particular application or center product layers to authorize validation and information uprightness. All passwords must be scrambled, and any associations with the framework ought to in a perfect world utilize Secure Sockets Layer/Transport Layer Security. Guarantee logs are produced from all exchanges around touchy information also.

2.3 Enormous information challenges in investigation

As said, to insert images in Word, position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with "Float over text" unchecked). The authors of the accepted manuscripts will be given a copyright form and the form should accompany your final submission. Conveyed programming system use parallelism in calculations and capacity to handle gigantic measures of the information. A prevalent illustration is guide diminish structure, which parts an information document into different pieces in the main period of delineate, a mapper for each lump peruses the information, play out some calculation, and yields a rundown of key/esteem pairs. In the following stage, a reducer joins the qualities having a place with each particular key and yields the outcome. There are two noteworthy assault aversion measures: securing the mappers and securing the information within the sight of an untrusted mapper. Non-social information stores advanced by NoSQL databases are as yet developing regarding security infrastructure. For case, hearty answers for NoSQL infusion are as yet not develop each NoSQL DBs were worked to handle distinctive challenges postured by the examination world and henceforth security was never some portion of the model anytime of its plan stage. Developers utilizing NoSQL databases normally install security in the middleware. Non-social information stores advanced by NoSQL databases are as yet developing as for security infrastructure. For case, vigorous answers for NoSQL infusion are as yet not develop each NoSQL DBs were worked to handle distinctive challenges postured by the examination world and thus security was never some portion of the model anytime of its outline stage. Developers utilizing NoSQL databases typically insert security in the middleware

2.4 Big data leakage

A noteworthy hazard in Big Data will be information spillage, which debilitates protection. Late discussions with respect to spilled records uncover the extent of huge information gathered and dissected over a wide range by the National Security Agency (NSA), also as other national security offices. This circumstance freely uncovered the hazardous harmony amongst security and the danger of sharp information abuse. In thought of security, the development of environment information may be influenced. Also, the adjustment of energy held by the government, organizations, and people has been exasperates, along these lines bringing about racial profiling and different types of disparity, criminalization, and constrained flexibility. In this way, appropriately adjusting pay dangers and the support of security in information is by and by the best test of open arrangement. In basic leadership with respect to significant strategies, staying away from this procedure initiates dynamic legitimate emergencies. Every accomplice addresses concerns with respect to security in an unexpected way.

For instance, common freedoms speak to the quest for outright power by the legislature. These freedoms fault protection for erotic entertainment and plane mischances. As per Birds of prey security, no favorable position is sufficiently convincing to balance the cost of awesome protection. Be that as it may, partners of information at no time in the future consider the danger of protection as they pursuit completely for data. Existing reviews on protection investigate the dangers postured by extensive scale information and gathering them into private, corporate, and administrative concerns; in any case, they neglect to recognize the advantages. Rubinstein proposed numerous structures to clear up the dangers of protection to choice creators and actuate action. As an outcome, business ventures what's more, the administration are progressively impacted by criticism concerning.

2.5 Big data on information secrecy

Information secrecy in big data most huge information specialist organizations or proprietors at present couldn't adequately keep up and break down such enormous datasets on account of their restricted limit. They should depend on experts or apparatuses to break down such information, which increment the potential security dangers. For instance, the value-based dataset for the most part incorporates an arrangement of finish working information to drive key business forms. Such information contains points of interest of the most minimal granularity and some delicate data such as charge card numbers. Consequently, investigation of enormous information might be conveyed to an outsider for preparing as it were at the point when legitimate preventive measures are taken to ensure such delicate information, to guarantee its security.

3. Related work

As per the overview of papers the new thought has been proposed on enormous information security. The information are put away in information hub. So the information hub is critical one. Before putting away the information the information hub must have security suite for security reason. It is act like a channel to diminish the mistake records. The unstructured information contains content, sound, video, and so on. The each document is put away safely with the assistance of the calculations quantum cryptography.

In this exploration, quantum cryptography furnishes greatest insurance with less intricacy that builds the capacity limit and security quality of the enormous information. In this segment, we have to recollect the utilization of symmetric key with a piece figure which is reasonable to control the enormous information security in light of the fact that the outline of the square figure for the huge information is extremely basic. Multifaceted nature dependably increments when we utilize vast squares yet we can limit the preparing steps progressively. Here, piece figure utilizing GA which gives proficient key hunt is one of the best QC approaches in huge information security strategies. Through this calculation, secure correspondences between the versatile clients and confirmation server can be set up. In symmetric key advancements, square figures outlined with GA are intense to make a productive key administration conspire for future server farms. Accept any piece figure of key size is n. Along these lines, taking after the condition decreases the means and multifaceted nature when mystery key is set up in the assault.

4. Flow of big data in security suite

In our approach we have fabricated a security suite to give required and sufficient security to information. The suite has four major segments with respect to security angles; first is for client distinguishing proof and verification which incorporates advanced signature plan or watchword confirmation plot, one is for privacy which contains encryption and decoding. The last one is for access control schemes. Each section is further divided into three sections which are for three classes of data according to sensitivity level. There is a scheduling algorithm which takes decision to activate appropriate security services from the selected section and provide adequate security according to the sensitivity level. The detail view of security suite has been shown in Figure 1. The calculations, the third is for respectability incorporates hash works, the fourth is for uprightness and confirmation, which incorporates MAC

Fig 1. View of Data node and Security suite

To manufacture the security suite we have utilized cover/code for each administration, for example, CS, HF and MC for the protection, uprightness, legitimacy with honesty individually. To give security to any information the framework gets to code related with it and select calculation from security suite. As for the information with code TXCS01 (see Table III), the calculation chooses 3DES is chosen to give cryptographic support of the information. It have considered measures or calculations of the diverse administrations as indicated by affectability levels of each kind of information.

For example, to keep up the security of content information in touchy class it have considered 3DES for information in delicate class. For the fundamental three administrations like security, uprightness and respectability with validation. For detail and ground of this sort of thought the peruser can be found in the papers [9-11]. To secure content sort information, Diffie-Hellman key trade conspire or advanced authentications can be utilized to comprehend key administration issue. For example, X.509 is a standard utilized as advanced authentication.

5. Conclusion

Along these lines the paper examined about the review in the huge information security and the procedures which is client to enhance the security. The new systems are touched base in the Hadoop outline work. The Hive, MapReduce devices are utilized to end up plainly more related security choices

References

[1] 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.

[2] Vijey Thayananthan and Aiiad Albeshri "Big data security issues based on quantum cryptography and privacy with authentication for mobile data center", *2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).*

[3] Md.Rafiqul Islam, Md.Ezazul Islam, "An approach to provide security to unstructured Big Data", Vol. 26, No. 1, January 2014.

[4] P.Kamaksh, "Survey on Big Data and Related Privacy Issues", *IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308.*

[5] Suren Kumar Sahu, Lambodar Jena, Santosh Satapathy, "Big Data Security issues and challenges in Cloud Computing Environment", *International Journal of Innovations in Engineering and Technology (IJIET).* vol. 10, no. 5, pp. 767-782, May 2001.

[6] Subaira.A.S, Gayathri.R, Sindhujaa.N, "Security Issues and Challenges in Big Data Analysis", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 6, Issue 2.

[7] Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani. "Big Data: Survey, Technologies, Opportunities, and Challenges", Volume 2014, *Article ID 712826.*

[8] Min Chen, Shiwen Mao, Yunhao Liu, "Big Data: A Survey", *Mobile Netw Appl. (2014) 19:171–209 DOI 10.1007/s11036-013-0489-0.*

[9] Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. *McKinsey Global Institute.* May 2011.

[10] "Big Data, Big Impact: New Possibilities for International Development*." World Economic Forum*.

[11] Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. *McKinsey Global Institute.* May 2011.

AUTHOR BIOGRAPHIES

Mr.S.Castro M.Tech is working as an Assistant Professor in the Department of Information Technology at Kalasalingam Institute of Technology, India. He is doing his Ph.D research in Anna university, Chennai. He published journal papers and presented papers in various conferences.

Dr.R.PushpaLakshmi Ph.D is working as an Associate Professor in the Department of Information Technology at PSNA College of Engineering and Technology, India. Her papers are published in various National, International journals and Conferences. Her area of research includes Wireless Networks, Network Security and Data Mining.