IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

Enhancing Information Security in Big Data

Renu Kesharwani

M.Tech Computer Science Department, United College of Engineering & Research, Allahabad, India

Abstract: Data Mining attracting more and more attention in recent years probably because of the growth of the "Big Data". This Big Data when analyzed properly can be very dangerous to insecure the information of Enterprise data. The biggest challenge for big data from security point of view is the protection of user's privacy. If the data is stored in the cloud a there should be establish a trust boundary between the data owners and the data storage owners. So the major challenge is to developing a safe and sound information sharing protocol .here we will discuss about the methods and privacy concern so that we could secure our user's sensitive information. Making use of Data Mining and some efficient algorithm user's sensitive privacy can be well taken .In this paper we will use the code inline parsing technique to make information more secure from the attackers and hackers and from the SQL injections so that we could make our Information more secure. This would be helpful to protect sensitive information on a big data platform.

Keywords: Information Security, Big Data Security, Big Data Analysis, Hadoop project.

I. INTRODUCTION

In 21st century the rapid advancement in field of information technology, Big Data has become very popular term in enterprises and industries. The growth of data expanded in very fast speed. This data comes from various sources form social media, digital picture, videos, purchase-transaction record, numbers etc. So managing such huge amount of data became more challenging job .These huge data is called as Big Data .This data can be processed and converted to some meaningful information and became very beneficial for the enterprises to understand the customer requirements by analyzing , collecting accurate data & more informed strategic decision on the basis of the analysis of Big Data But from the security point of view Big Data increases opportunity for attackers they insert malicious software in apps and operating systems. So the aim here is to implement some enhanced data security mechanism for making privacy of users data more secure.

II. GROWTH OF BIG DATA

An easy way the data is increasing day by day, we are familiar with the websites like facebook, twitter, and all the users are entering the data at the same time. According to IBM 90% of the data in the world has been created in last two years alone and every day we create 2.5 quintillions bytes of data [2] .The biggest problem with the Big Data is how to store the huge amount data?

So for this the large database are needed that cannot be managed efficiently by common database management system. These datasets range from terabytes to exa bytes and the other challenge is to secure these data so that the privacy & protection of user's data can be maintained. The process of storing and analyzing the huge data to make sense for betterment of organization is "Big Data". In this paper we will discuss about the growth of data and how can we enhance the security of the data.

III.MEASURE SECURITY ISSUES

All The major challenge for the Big Data from security point of view is to secure the user's information privacy from unauthorized users. For example when we enter any information that store in the databases and some hackers or attackers try to access those information from the databases.

There is various security issues by which the information security are breaching.
SQL injection. Attackers and hackers passes the set of codes which are called SQL injection to break the database access. They use some default codes which helps them to break the security.
Big Data Skill Gap. As data analytics become broad to businesses around the globe; many companies are facing a ability shortage for these skills. This skill gap has held to more companies moving to online training to get both new and old workers up to speed quickly.
So there are various ways who breaches the security. User identification and authentication are the major aspects of confidentiality. Who gets to access what information also needs to be considered here. Hackers and unauthorized user activity, infected files when downloaded and virus attacks are some of the most commonly encountered threats to information confidentiality.

IV. APPROACH TO THE TECHNIQUE OF BIG DATA SECURITY

After looking at the broader prospect many businesses across the globe use the big data for marketing and research yet may not have fundamentally right looking at the security prospect, like the general trend with new securities comes at the last in the list causing a major concern. Breaches with big data is similarly immense like the technology, its potential is even more serious reputational damage and legal repercussions than at present. With increase in the use of big data to store and analyze

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

petabytes of data including weblogs, social media content and click stream data to gain the better insight of customer's capacity and their business .It results into more classified information its ownership is facilitated any classification reasonably.

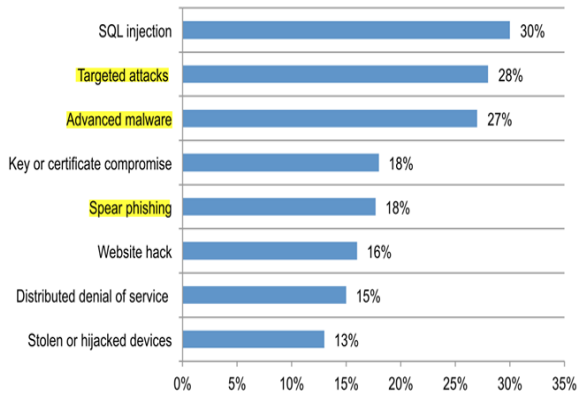

Fig 1.How the malicious or criminal breach

Most organizations already struggle with implementing these concepts, making this a significant challenge. For this we have to identify for outputs of the owners and their raw data of Big Data Processes. Thus data ownership will be distinct from information ownership – perhaps with IT owning the raw data and business units taking responsibility for the outputs. Very few organizations are likely to build a Big Data environment in-house, so cloud and Big Data will be inextricably linked. Nowadays, organizations are collecting and processing massive amounts of information. The more data is stored, the more vital it is to ensure its security. A lacking of data security can lead to large financial losses and reputational damage for a company. As far as Big Data is concerned, losses due to lack of IT security can exceed even the worst expectations. As many enterprises are aware, storing data in the cloud does not remove their responsibility for protecting it - from both a regulatory and a commercial perspective. Increasing risks of cybercrime and other malicious activity on the Internet is prompting enterprises to deploy more security controls and collect more data than ever before. As a result, advances in big data analytics are now being applied to security monitoring for broader and more in-depth analysis to protect valuable company resources.

V. PROPOSED SOLUTION

A. Technical solution to privacy protection

To enhance the security we will take the input from user then we will add the prefix and suffix to encrypt the original input data so that the encrypted data input could be save to the database table and the original data will be save to other table. By storing data into two different tables we can enhance the security. When attacker will try to access the database he will find the encrypted database in place of original database. So the security will be enhanced.

In Present Scenario the frontend directly connected to the database via single layer database connectivity when attacker or malicious software passes the Database can be drop easily.

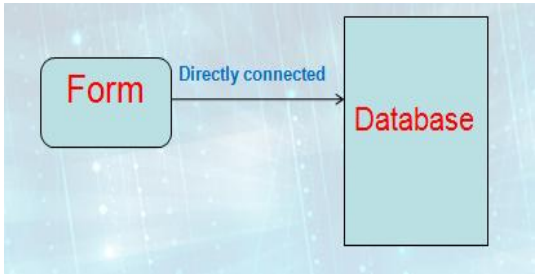

Fig 2: Simple Database connection technique

Here query will be pass via two functions that will be interrelated and trim and encrypted result will be stored to one database and main data will be secure to the another database.

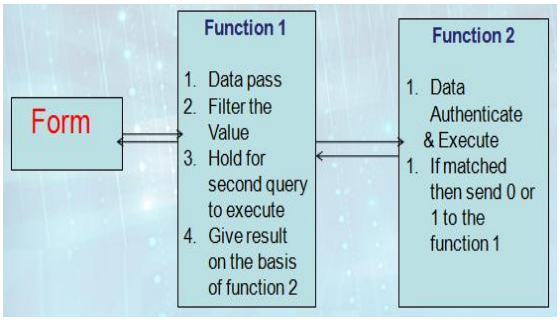

Fig 3: Enhanced Database connection technique

B. Non Technical solution to privacy protection

In above section we mainly explore the technique and approaches to the privacy issues in security but some security issues incidents remind us for the non technical solutions such as laws, industrial deals and regulations are also responsible for the security issues of user's privacy [3]. User's information privacy protection is major responsibility for legislation. Many countries have their laws for securing the users privacy so that their personal information can be secured but still there is need to improve current legislation to reconcile the laws for the individual's rights to privacy and the government's authority law for accessing personal information for national security. Also it is necessary for awareness of public to provide them education for the information security.

VI. BIG DATA SECURITY CHALLENGES

In this paper, we will firstly discuss the enterprises benefits and challenges of Big Data security and privacy .Then will explain some techniques and solution ensure security and privacy in Big Data. Attackers who are using these detecting and blocking advanced persistent threat techniques may employ slow-paced, low-visibility attack patterns to avoid detection [17].

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

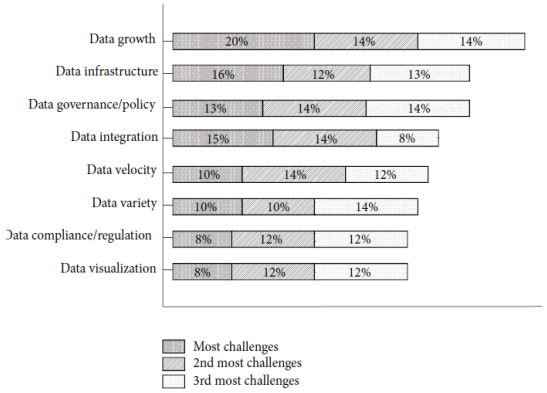

Fig 4: Challenges in Big Data

This is the approach used for analytics platforms in Big Data security [13].

- Unethical IT specialists practicing information mining can gather personal data without asking users for permission or notifying them.
- Non-relational databases (NoSQL) are actively evolving, making it difficult for security solutions to keep up with demand.
- For Most distributed systems' computations have only a single level of protection, which is not recommended.
- Recommended detailed audits are not routinely performed on Big Data due to the huge amount of information involved.
- Recommended detailed audits are not routinely performed on Big Data due to the huge amount of information involved.
- Due to the size of Big Data, its origins are not consistently monitored and tracked.
- Some organizations cannot – or do not – institute access controls to divide the level of confidentiality within the company.
- When a system receives a large amount of information, it should be validated to remain trustworthy and accurate; this practice doesn't always occur, however.
- Access control encryption and connections security can become dated and inaccessible to the IT specialists who rely on it.
- Automated data transfer requires additional security measures, which are often not available.

VII. APPLICATION AREAS OF BIG DATA

A. Figure Big Data for Healthcare

In healthcare, the pharmacy and medical device companies use Big Data to improve their research and development practices.. Big Data also helps researchers to work toward eliminating health-care related challenges before they become real problems [12]. Big Data helps doctors to analyze the requirements and medical history of every patient and provide individualistic services to them, depending on their medical condition.

Traditional health data centers capture and store an huge amount of structured data which included concerning a wide range of information including diagnostics, laboratory tests, medication, and ancillary clinical data.. Thus, advances in big data processing for health informatics, bioinformatics, sensing, and imaging will have a great impact on future clinical research.

A. Government

Big Data has come to play an important role in almost all the undertaking and processes of government. In the field of governance challenges also raises in terms of privacy, security, data stewardship and data ownership.

- By analyzing data help to identify flaws and loopholes in process.
- It helps to taking informed decisions in time about various issues.
- Preventing fraudulent practices in various sectors.
- Improving Various sectors such as education , health, defense and research areas
- Helps to invest budget in beneficial areas.

B. Consumer Goods Industry

Consumer goods companies generate huge volume of data in varied formats from different sources, such as transaction, billing details, feedback forms etc [14]. This data should be organised and analyzed in a systematic manner to generate information which helps enterprises and industries to achieve profits and prevent from losses to the company .Therefore we can say that Big Data analytics allows organisation to gain better business insights and take informed and timely decisions.

VIII. ADDITIONAL SOLUTIONS FOR BIG DATA SECURITY ENHANCEMENT

Solution by tapping into new volumes and varieties of data, scientists, executives, product managers, marketers, and a range of others can have plans and decisions to discover new opportunities for optimization, and deliver breakthrough innovations.

• Analytics. The output is ultimate fruit of a big data, the analytics that help the business optimize and innovate. In database the information are stored can be presented in dashboards and reports, and can be accessed via on-demand queries. Big data analytics represent the most sensitive asset of all in some businesses, intelligence that provides a critical competitive differentiator—and a huge competitive exposure if it falls into the wrong hands.

• Data sources. To most fully exploit the advantages of big data, organizations leverage various forms of data, including both structured data in a range of heterogeneous applications and databases and unstructured data that come in a number of file types. Organizations may leverage data from enterprise resource planning systems, customer relationship management platforms, video files, spreadsheets, social media feeds, and many other sources. Further, more data sources are added all the time. Today,

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

you don't know where new data sources may come from tomorrow, but you can have some certainty that there will be more to contend with and more diversity to accommodate. These big data sources can include personally identifiable information, payment card data, intellectual property, health records, and much more. Consequently, the data sources being compiled need to be secured in order to address security policies and compliance mandates.

• Big data frameworks. Within the big data environment itself—whether it's powered by Hadoop, MongoDB, NoSQL, Teradata, or another system—massive amounts of sensitive data may be managed at any given time. Sensitive assets don't just reside on big data nodes, but they can come in the form of system logs, configuration files, error logs, and more.

In big data environments, data is routinely replicated and migrated among a large number of nodes. In addition, sensitive information can be stored in system logs, configuration files, disk caches, error logs, and so on. Vormetric Transparent Encryption efficiently protects data across all these areas, delivering encryption, privileged user access control, and security intelligence. In addition, with Vormetric Protection for Teradata Database, your organization can gain the comprehensive, granular controls required to secure the most sensitive assets across your Teradata environments, while enabling you to maximize the business benefits of your big data investments. Cloud computing experts believe that the most reasonable way to improve the security of Big Data is through the continual expansion of the antivirus industry.[16] A multitude of antivirus vendors, offering a variety of solutions, provides a better defense against Big Data security threats. Refreshingly, the antivirus industry is often touted for its openness. Antivirus software providers freely exchange information about current Big Data security threats, and industry leaders often work together to cope with new malicious software attacks, providing maximum gains in Big Data security.

Here are some additional recommendations for Big Data security improvement:

- We should Focus on application security, than into the device security.
- Isolate devices and servers containing critical data.
- Work on Providing reactive and proactive protection.
- Introduce real-time security information and event management.

Immediate concern to companies using Big Data is the security of cloud-based systems. Intel Security has recently published the McAfee Labs' Threat Predictions Report that contains their expectations for the near-future of data security. Of particular concern in this report is the supposition that legitimate cloud files hosting services such as Dropbox, Box, and Stream Nation are at risk of being used as control servers in upcoming cyber espionage campaigns. If targeted, these popular cloud services could enable the malware to transfer commands without raising suspicion. Malicious attacks on IT systems are becoming more complex and new malware is constantly being developed. Unfortunately, companies that work with Big Data face these issues on a daily basis. Nevertheless, every problem has a solution and finding an effective and suitable answer for your organization is indeed possible.

IX. FUTURE SCOPE

You must Today, Big Data is influencing IT industry like few technologies have done before. The massive data generated from sensor-enabled machines, mobile devices, cloud computing, social media, satellites help different organizations improve their decision making and take their business to another level.

"Big data has the potential to change various organizations, academic institutions and governments conduct business and make discoveries and it is likely to change how everyone lives their day-to-day lives," - Susan Hauser, corporate vice president of Microsoft. Data is the biggest thing to hit the industry since PC was invented by Steve Jobs.[16] As mentioned earlier in this paper, every day data is generated in such a rapid manner that, traditional database and other data storing system will gradually give up in storing, retrieving, and finding relationships among data. Big data technologies have addressed the problems related to this new big data revolution through the use of commodity hardware and distribution. Companies like Google, Yahoo!, General Electric, Cornerstone, Microsoft, Kaggle, Facebook, Amazon that are investing a lot in Big Data research and projects. IDC estimated the value of Big Data market to be ―about $ 6.8 billion in 2012 growing almost 40 percent every year to $17 billion by 2015.[16] By 2017, Wikibon's Jeff Kelly predicts the Big Data market will top $50 billion. All companies are exploring big data strategies because demand is so hot for solutions.

The companies problem is lack of internal expertise and good practices. It's a perfect storm of product and services says Wikibon's Jeff Kelly. Recently it was announced that, Prime Minister's of Indians office is using Big Data analytics to understand Indian citizen's sentiments and ideas through people oriented crowd sourcing platform www.mygov.in and social media to get a picture of common people's thought and opinion on government actions.

Google is launching the Google Cloud Platform, which provides developers to develop a range of products from simple websites to complex applications. It enables users to launch virtual machines, store huge amount of data online, and plenty of other things. Basically, it will be an one stop platform for cloud based applications, online gaming, mobile applications, etc. All these required huge amount of data processing where Big Data plays an immense role in data processing. [15] In Future I will

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

definitely want to research more in Big Data security field, here in this paper I am using code parsing technique to make database more secure. in future I will find method to create two databases where in one database main record will be stored and another the encrypted record will be saved and only on the basis of reference of main data databases will work and after limited time of period the reference record would be removed so that security could be more increases.

X. CONCLUSION

Big data increases opportunity for attackers they Insert malicious software in apps and operating systems and use various attacking techniques to get accessing into the database where huge amount of data are being stored. In this paper the purpose is to use the Big Data techniques for enhancing the security of user's privacy To Write this paper my aim is to analyze the Big Data so that it could be easy to find out what major issues are occurring, what kind of benefits it could provide to the enterprises and how it can harm to user's privacy and security.

REFERENCES

[1] G Geethakumari "Big Data Analysis for Implementation of Enterprise Data Security ", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 2, No.4, August 2012
[2] Cloud Security Alliance " Big Data Analytics for Security Intelligence" September 2013
[3] LEI XU, CHUNXIAO JIANG, (Member, IEEE), JIAN WANG, (Member, IEEE)" Information Security in Big Data: Privacy and Data Mining", Received September 21, 2014, accepted October 4, 2014, date of publication October 9, 2014, date of current version October 20, 2014.
[4] Mr. Mohammad Raziuddin & Prof. T.Venkata Ramana "Literature Survey in Data Mining with Big Data" International Journal of Advanced Engineering and Global Technology I Vol-03, Issue-04, April 2015.
[5] Review Article Big Data: Survey, Technologies, Opportunities, and Challenges , Volume 2014 http://dx.doi.org/10.1155/2014/712826
[6] Roger Schell "Security – A Big Question for Big Data" 2013 IEEE International Conference on Big Data.
[7] IBM "Big Data at the speed of Business", "http://www-01.ibm.com/software/data/bigdata/2012.
[8] http://bigdataarchitecture.com/
[9] http://en.wikipedia.org/wiki/Apache_ Hadoop
[10] "Big Data is the Future of Healthcare Bill Hamilton" cognizant 20-20 insights 2012
[11] Sam curry "big data fuels intelligence –driven security" RSA Security brief 2013
[12] "Big Data for Health" Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang, Fellow, IEEE
[13] http://www.datacenterknowledge.com/archives/2016/01/19/nine-main-challenges-big-data-security/
[14] Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization
[15] "Big Data A New World of Opportunities", NESSI White Paper, December 2012.
[16] Samiddha Mukherjee, Ravi Shaw" Big Data Concepts, Applications, Challenges and Future Scope" Vol. 5, Issue 2 , February 2016
[17] CLOUD SECURITY ALLIANCE Expanded Top Ten Big Data Security and Privacy Challenges, April 2013

BIOGRAPHY

Renu Kesharwani is a master student (M.Tech) in the Department of Computer Science and Technology, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh, Lucknow, India. She received her M.Sc. Computer Science degree (Gold Medalist) from APS University of Rewa India in 2013. Her research interests include Data Mining, Information Security and big data security.