Securing Big Data From Cyber Threats

M.Mounika, N.Vijaya lakshmi, Dr.K. Venkata Ramana,principal Dept.of MCA

Abstract— Big data presents a challenge for data base and data analytical research. Big data provides the unstructured data. Big data will also help analysts to visualize cyber attacks by taking the complexityfrom various data sources and security solutions to protect data and prevent future cyber *attacksTerm* Big Data Analytics for Security intelligence refers to a process of analyzing and mining large amounts of data (petabytes, exabytes, zettabytes) from different sources including IP address, Emails, log files, information get from other attack investigation and many more.

1.INTRODUCTION :

The age of big data and cyber security is here. And that means both opportunity and risk for most businesses. If you are in the cyber security field you are likely very familiar with big data, which is the term used to describe a very large data set that is mined and analyzed to find patterns and behavioral trends. It is generally defined [1]as being dense in variety, velocity and volume. From a cyber security standpoint big data has ushered in new possibilities in terms of analytics and security solutions to protect data and prevent future cyber attacks. But just as big data has opened up new possibilities for cyber security teams, it has also given cyber criminals the opportunity to access mass quantities of sensitive and personal information through the use of advanced technologies.

M.Mounika, Dept. of MCA 2nd year, KMMIPS, Tirupati, mail id:mounikam137@gamil.com.

N.Vijaya lakshmi, Dept. of MCA 2nd year, KMMIPS, Tirupati, mail id:vijayalakshmireddy369@gamil.com.

Dr.K. Venkata Ramana,Principal Dept .of MCA, KMMIPS, Tirupati, mail id: ramanakv4@gmail.com.

1.1 BIG DATA :

Big data is similar to the small data but it is in bigger size.The main aim of big data is to solve new problems or old problems in a better way.It generates value from the storage and processing of very large quantities of digital information.Big data cannot be analyzed with traditional computing techniques.Big data requires a set of techniques and technologies.Big data has to help companies improve operations and make faster,more intelligent decisions.

1.2 Applications of bigdata:

Big Data is slowly becoming ubiquitous. Every arena of business, health or general living standards now can implement big data analytics. To put simply, Big Data is a field which can be used in any zone whatsoever given that this large quantity of data can be harnessed to one's advantage.

Figure1: Applications of Big Data

1) IN BDA: Big Data Analytics Applications (BDA Apps) are a new type of software applications, which analyse big data using massive parallel processing frameworks (e.g., Hadoop). Developers of such applications typically develop them using a small sample of data in a pseudo-cloud environment. Afterwards, they deploy the applications in a large-scale cloud environment with considerably more processing power and larger input data (reminiscent of the mainframe days) Big Data Analytics Applications (BDA Apps) are a new category of software applications that leverage largescale data, which is typically too large to fit in memory or even on one hard drive, to uncover actionable knowledge using large scale parallel-processing infrastructures [1]. The big data can come from sources such as runtime information about traffic, tweets during the Olympic Games, stock market updates, usage information of an online game [2], or the data from any other rapidly growing data-intensive software system.

2) In Data Mining: Decision Tree--Datameer's decision trees automatically help users understand what combination of data attributes result in a desired outcome. Decision trees illustrate the strengths of relationships and dependencies within

data and are often used to determine what common attributes influence outcomes such as disease risk, fraud risk, purchases and online signups. The structure of the decision tree reflects the structure that is possibly hidden in your data.

3) In Banking: The use of customer data invariably raises privacy issues. By uncovering hidden connections between seemingly unrelated pieces of data, big data analytics could potentially reveal sensitive personal information. Research indicates that 62% of bankers are cautious in their use of big data due to privacy issues. Further, outsourcing of data analysis activities or distribution of customer data across departments for the generation of richer insights also amplifies security risks. For instance, a recent security breach at a leading based bank exposed databases of thousands of customer files. Although this bank launched an urgent investigation, files containing highly sensitive information. Such as customers' earnings, savings, mortgages, and insurance policies ended up in the wrong hands10. Such incidents reinforce concerns about data privacy and discourage customers from sharing personal information in exchange for customized offers.

4) In Enterprise: For enterprises around the world, in many industries, in-database analytics are providing a competitive advantage. When data doesn't have to commute to work and back, it can deliver faster insights that help businesspeople make informed decisions in real time for less expense than traditional data analysis tools.

5) In Agriculture: A biotechnology firm uses sensor data to optimize crop efficiency. It plants test crops and runs simulations to measure how plants react to various changes in condition. Its data environment constantly adjusts to changes in the attributes of various data it collects, including temperature, water levels, soil composition, growth, output, and gene sequencing of each plant in the test bed. These simulations allow it to discover the
optimal environmental conditions for specific gene types.

6) In Finance: A major financial institution grew wary of using third-party credit scoring when evaluating new credit applications. Today the bank performs its own credit score analysis for existing customers using a wide range of data, including checking, savings, credit cards, mortgages, and investment dat.

7) In Economy: Designed from the ground up to deal intelligently with commodity hardware, Hadoop can help organizations transition to low-cost servers.

8) In Consumer Goods: A maker of consumer products collects consumer preference and purchasing data extracted from surveys, purchases, web logs, product reviews from online retailers, phone conversations with customer call centres, even raw text picked up from around the Web. Their ambitious goal: to collect everything being said and communicated publicly about their products and extract meaning from it. By doing this, the company develops a nuanced understanding of why certain products succeed and why others fail. They can spot trends that can help them feature the right products in the right marketing media.

9) Where to Use Hadoop: In every vertical there are data tasks with which Hadoop can assist. These tasks have different terms depending on the industry but they all come down to either advanced analytics or data processing.

10) In Smart Phones: Perhaps more impressive, people now carry facial recognition technology in their pockets. Users of I Phone and Android smart phones have applications at their fingertips that use facial recognition technology for various tasks. For example, Android users with the remember app, can snap a photo of someone, then bring up stored information about that person based on their image when their own memory lets them down a potential boon for salespeople. I Phone users can unlock their device with recognize me, an app that uses facial recognition in lieu of a password. If deployed across a large enterprise, this app could save an average of $2.5 million a year in help-desk costs for handling forgotten passwords.

1.3 SECURING FOR BIG DATA:

Almost every big data security issue that's common to enterprise-wide implementations can be traced back to design omissions in the original Hadoop distribution. Not that Hadoop's original design was faulty or bad, it just was not designed to be used in an enterprise data environment. Enough time has passed, however, that effective adaptations and solutions have been developed to address these security concerns. The trick is to first understand what the potential weaknesses are, and

then verify that you have taken proper precautions to protect those weaknesses.

- **User Authentication and Access** – Organizations deploying Hadoop in a shared environment must be sure that user authentication and access rights are strictly controlled. Apache Sentry is one possible solution that's available to help you limit and control user access rights across a big data system.
- **Regulatory Requirements** – With so much data, organizations have to make a real and concerted effort to comply with regulatory requirements. In big data systems, you'd be prudent to take a few steps of extra precaution by ensuring that records on user.
- activities and system events are being generated and stored. You'll need them to carry out any user and system audits.
- **User Impersonation** – Several big data security issues center around the fact that user and service authentication protocols in native Hadoop are somewhat weak. This leaves Hadoop systems open to the risk of malicious data inputs and edits. Make sure you have Kerberos and LDAP protocols in place in order to safeguard against this weakness.
- **Protecting Data-At-Rest and Moving Data** – Native Hadoop distributions offer data encryption capabilities for data-at-rest, but it's a bit trickier to protect data-in-motion. Network encryption methods have been developed to protect moving data, but they're not included with Hadoop's native distribution, so you'll need to set up that line of protection for yourself.

2. SECURITY THREAT TO BIG DATA:

In this section, we present big data security lifecycle model and the main components of any big data framework. We extend our model from (Xu et al. 2014). They address big data from user role perspective where they argue four types of users' role in big data environment: data provider, data collector, data miner, and decision maker. However, our model addresses the phases of the big data lifecycle. Our model consists of four phases in big data framework consists of data collection phase, data storage phase, data processing and analysis, and knowledge creation. Figure 1 presents the main elements in big data lifecycle.

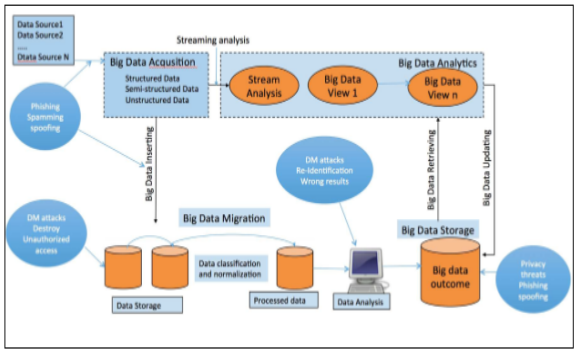

Figure2: Big Data Lifecycle Threat

Data collection phase : In data collection phase, data from different sources comes with different formats: structured, semistructured, and unstructured. From a security perspective, securing big data technology should start from the first phase of the lifecycle. It is important to gather data from trusted sources and make sure that this phase is secured and protected. In fact, we need to take some security measures in order to keep data from being released. Some security measures can be used in this phase like limited access control (for those who receive data from data provider) and encrypting some data fields (personal information identifier).

Data storage phase: In data storage phase, the collected data is stored and prepared for being used in the next phase (data analytics phase). As the collected data may contain of sensitive information, it is essential to take sufficient precautions during data storing. In order to guarantee the safety of the collected data, some security measures can be used like data anonymization approach, permutation, and data partitioning (vertically or horizontally).

Data Analytices phase: After collecting data and storing it in secured storage solutions, data processing analysis is performed to generate useful knowledge. In this phase, data mining methods such as clustering, classification, and association rule mining are used. It is crucial to provide secure processing environment. In fact, data miners use powerful data mining algorithms that can extract sensitive data. Thus, a security breach may happen. Therefore, data mining process and it's output must be protected against data mining

based attacks and make sure that only authorized staff work in this phase.

Knowledge Creations Phase: Finally, the analytics phase comes up with new information and valued knowledge to be used by decision makers. The created knowledge is considered as sensitive information especially in a competition environment. Organizations take care of their sensitive information to be far away from their rivals. Further, they aware of their sensitive data (e.g. client personal data) not to be publicly released.

3.SOLUTIONS OF BIG DATA SECURITY:

Cyber security needs the risk management and actionable intelligence that is common from big data analysis. While it is great to have tools that can analyze data, the key is to automate tasks so that the data is available more quickly and the analysis is sent to the right people on time. This will allow analysts to classify and categorize cyber threats without the long delays that could make the data irrelevant to the attack at hand.Big data will also help analysts to visualize cyber attacks by taking the complexity from various data **sources** and simplifying the patterns into visualizations.Being able to utilize the data in its raw format allows disparate data to be useful not only with what is happening now, but also with historical data. Using this historical data, you can create statistical baselines to identify what is "normal." You will then be able to determine when the data deviates from the norm. Sometimes it's easy to miss indicators when they are offered in real time; however, they may have new meaning when they are viewed over time.

This historical data can also create new possibilities for predictive models, statistical models, and machine learning. This gives the ability to predict future events. However, it's what you can do with this data, if anything, that can make the difference between being attacked or not. After all, data is just really information unless an action is taken towards improving cyber security. Being able to automatically respond to threats noticed in data, and also being able to have a high level of trust in the accuracy of the data is key to a big data security solution.

4. Conclusion

Some might believe that big data will quickly solve the problems of the cyber security industry. The reality is that data and analytics will allow companies to identify anomalies and advanced attack vectors. uses machine learning paired with cloud intelligence and automated responses to detect unusual activity and respond when you need it.

The main aim of this paper is to explore the role of Big Data in various fields. Big Data is a very powerful tool that makes things ease in various fields as said above. Big data used in so many applications they are enterprise, ecanamy, banking, agriculture, chemistry, data mining, finance, marketing, stocks, BDA, health care etc… We can see in this paper various fields and use the big data application.

5.Referneces

[1]https://onlinedegrees.sandiego.edu/threat-or-opportunity-big-data-and-cyber-security/

[2]Sokolva M, Matwin s.personal privacy protection in time of Big Data. Berlin: Springer; 2015

[3] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," interactions, vol. 19, no. 3, pp. 50–59, May 2012

[4] N. Wingfield, "Virtual product, real profits: Players spend on zynga's games, but quality turns some off," Wall Street Journal.

[5]Tsai C-W, Lai C-F, Chao H-C, Vasilakos AV. Big data analytics: a survey. J Big Data Springer Open J. 2015.

[6].http://www.ibm.com/software/data/bigdata/industry-healthcare.html.

[7]http://www.firstpost.com/business/big-data-booster-shothealthcare-industry-needs-2160271.html.

[8]http://blogs.worldbank.org/voices/meet-winners-and-finalists-firstwbg-big-data-innovation-challenge.

[9] Ari Banerjee senior analylist, heavy reading, "Big data and advanced analytics in Telecom: A Multi-Billion-Dollar Revenue Opportunity," December 2013.

[10] Weiyi Shangy, Zhen Ming Jiangy, Hadi Hemmatiy, Bram Adamsz, Ahmed E. Hassany, Patrick Martinx, "Assisting Developers of Big Data Analytics Applications

When Deploying on Hadoop Clouds" Database Systems Laboratory, School of Computing, Queen's University, Kingston, Canada.