International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Risk mitigation strategies of Big data with IOT

Vani Krishnaswamy
Assistant Professor,
Department of Decision and Computing Sciences,
Coimbatore Institute of Technology
vanithanu@cit.edu.in

Abstract

Today,technological advances are gaining momentum in the lives of users,but also in the world of business,health,industry and the military.One of the most promising technologies is the IoT(Internet of Things) which will allow physical objects to connect to the Internet, thus optimizing their functioning by generating data. However, in a world where data is becoming king, it must be handled efficiently and the means of IT must allow to store an ever-increasing number of data. This is where Big Data takes on its importance.The efficient mining of Big Data enables to improve the competitive advantage of companies and to add value for many social and economic sectors. In fact, important projects with huge investments were launched by several governments to extract the maximum benefit from Big Data and also private sector deployed important efforts to maximize profits and optimize resources. However, Big Data sharing brings new information security and privacy issues. Traditional technologies and methods are no longer appropriate and lack of performance when applied in Big Data context. This chapter presents Big Data security challenges and a state of the art in methods, mechanisms and solutions used to protect data-intensive information systems.

Keyword: Big Data, IoT,Security, Privacy, Data Security, Data Privacy.

1. Introduction

The rapid growth of global data by both individuals and corporations is partially attributed to the unexpected rise of unstructured data such as photos, videos and generally what social media has introduced to us and is expected to continue by a dramatic increase rate of 4300% in annual data generation by 2025 making data production 44 times greater in the year 2020 in comparison to 2009 . Increase data production in accordance with recent advances in storage technologies (such as cloud) has led to capture and storage of huge amounts of data called Big Data by academics, media and within the industry , which can be described as huge data sets with a variety of data types and a high velocity of streaming based on a report by Gartner Group.

Big data is gaining more and more attention since the number of devices connected to the so-called “Internet of Things” (IoT) is still increasing to unforeseen levels, producing large amounts of data which needs to be transformed into valuable information. Additionally, it is very popular to buy on-demand additional computing power and storage from public cloud providers to perform intensive data-parallel processing. In this way, security and privacy issues can be potentially boosted by the volume, variety, and wide area deployment of the system infrastructure to support Big Data applications.

Data analytics is being used in our everyday lives for extraction of patterns and knowledge from huge datasets providing businesses with new paradigms and governments with enhancement of their authorities . Few examples will include eBay.com, which has implemented a Hadoop cluster to improve its recommendation engine, or Facebook and Twitter storing queries for further analysis using data mining techniques. Another example would be Barack Obama’s 2012 re-election, during which, Big Data analytics were used for accurately discovering and addressing the political interest of the voters . Traditional mechanisms and policies are unable to address the security and privacy issues facing Big Data in today’s

computational environment; therefore, there is a need to re-visit issues like distributed environments, encryption algorithms, data storage, and real-time monitoring. In this paper, we thoroughly examine some of the root causes contributing to security and privacy breaches in Big Data to gain a better understanding of important research areas that should be given high priority when considering development of new methods. Section II explains briefly on Big Data definition and characteristics, while section III explains about IoT and its relation with Big Data. Section IV categorize, and investigates the main security and privacy concerns in relation to Big Data within current literature. In Section V, we further analyse how Big Data can be utilized to maintain security and privacy, and finally, in last section conclusion provides an overview of important topics discussed, and necessary requirements to secure Big Data communication.

2. Big Data Definition

The term Big Data is normally used for large and complex datasets that cannot be processed/managed by typical software which is characterized via 5Vs namely as volume (data size), velocity (high speed of data), variety (diverse data types and sources), veracity (consistency and trustworthiness of data), and value (outputs gained from data set) Figure 1 shows the different characters of Big Data via 5Vs.

A. **Volume:**

The capability of processing large amounts of data is a critical aspect of Big Data especially since volume is one of the biggest challenges of conventional IT structures in which companies are unable to process their large amounts of archived data logs. One example of such businesses is WalMart, which used to store 1,000 terabytes of data in 1999 as opposed to over 2.5 petabytes of data in the year 2012.

B. **Velocity:**

This points to the high speed at which data is created, processed, stored, and analysed by relational database in addition to the speed at which new data is generated and moved around like the way information on social media goes viral in matter of seconds or the hundred hours of video content uploaded to YouTube daily.

C. **Variety:**

Variety is another interesting aspect of Big Data, meaning that this data can come in structured, unstructured, or semi-structured form, making it extremely challenging for placement in a relational database, especially since in 90% of cases, the generated data is in unstructured form, making it crucial for data analysts to know the category to which Big Data belongs.

D. **Veracity:**

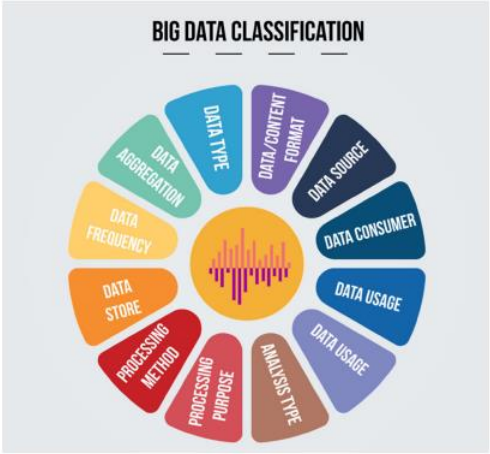
When dealing with Big Data, there is always the possibility of receiving dirty data (which is not 100% correct). The data quality and accuracy of analysis largely depends on the veracity of data source.

E. Value:

Even though there are great potential values in usage of Big Data unless there is a return on investment (value generated) for the company; it would be very costly (and useless) to

implement IT infrastructure systems to store Big Data.

We can use different approaches to acquire, process, store and analyse Big Data; however, it is important to keep in mind that there are different characteristics to sources from which Big Data is received, such as data type, size, speed, consistency/trustworthiness and frequency. Additionally, selection and built of a Big Data solution can be challenging due to factors like governance, security, and policies. Big Data can be categorized per the following classifications: data type, content, source, consumer, usage, analysis type, processing purpose, processing method, store, and frequency as illustrated in Figure 2 below



3.IoT Definition

The Internet of Things (IoT) is a concept that connects physical or virtual objects to the internet. The technology very often used is the sensor, allowing to link a physical object such as a watch, a drone or even a speaker, to the internet. If for a long time the few objects connected to the Internet were the telephone and the computer, this is no longer the case today and every year new types of objects incorporating IOT technology are born.

Each IOT has 5 common and inseparable components for the functionality of it. These are:

- **Sensors:** connecting the physical object to different computer systems;
- **Connectivity**: the network is essential to connect the object to the Internet (Wifi, wired, 4G or soon 5G...);
- **Data:** The main purpose of IOT is to collect and transmit data;
- **Information:** Translating data into information is essential to be able to read and then exploit the data;
- **Operating applications**: allowing you to control IOTs but also to read the information you receive.

IOT is one of the greatest technological revolutions of our era and its potential for exploitation is immense. IOT could have a huge impact on the cars of the future or on the new versions of **smart-cities**, an urban space connected to the Internet, thus significantly improving the lives of users, while reducing the negative impact of these on the planet.

4. Relationship between Big Data and IOT

According to several studies, the use of IOT is expected to generate **4.4 trillion GBytes in 2025,** and this figure is expected to increase in subsequent years. In addition, this data must be read, exploited and transmitted within specific timeslots, so, as you might have guessed, the major challenge in the field of the Internet of Things is to be able to exploit a huge amount of data, hence the use of big data.

TheRole of Big Data in IoT

Big Data should enable **real-time analysis of the data generated by IOT** and thus optimize the use of this technology. To do this, Big Data proceeds in four steps:

1. Collecting data generated by IoT by following the three primary principles of Big Data: speed, volume and variety.
2. Storing data in files within the Big Data database.

3. Data analysis by complex and efficient analytical systems, such as Spark or Hadoop
4. The implementation of the report of the analyzed data.

Big Data will play an important role in information processing efficiency and will enable IoT developers to optimize these tools to broaden the current perspective.
The interaction between IoT and Big Data is not one-way. IoT could also bring a lot to Big Data. The more important IoT are in your daily life and that of your city, the more developers will be demanding greater capacity in terms of big data and the more this business will grow.

It will thereby be important to improve data storage technologies to develop systems capable of **processing even more data.** This interaction could thus enable technological growth in both areas simultaneously.

5. Big data challenges to information security and privacy

With the proliferation of devices connected to the Internet and connected to each other, the volume of data collected, stored, and processed is increasing everyday, which also brings new challenges in terms of the information security. In fact, the currently used security mechanisms such as firewalls and DMZs cannot be used in the Big Data infrastructure because the security mechanisms should be stretched out of the perimeter of the organization's network to fulfill the user/data mobility requirements and the policies of BYOD (Bring Your Own Device). Considering these new scenarios, the pertinent question is what security and privacy policies and technologies are more adequate to fulfill the current top Big Data privacy and security demands (Cloud Security Alliance, 2013). These challenges may be organized into four Big Data aspects such as infrastructure security (e.g. secure distributed computations using MapReduce), data privacy (e.g. data mining that preserves privacy/granular access), data management (e.g. secure data provenance and storage) and, integrity and reactive security (e.g. real time monitoring of anomalies and attacks). Considering Big Data there is a set of risk areas that need to be considered. These include the information lifecycle (provenance, ownership and classification of data), the data creation and collection process, and the lack of security procedures. Ultimately, the Big Data security objectives are no different from any other data types – to preserve its confidentiality, integrity and availability.
Cloud Secure Alliance (CSA), a non-profit organization with a mission to promote the use of best practices for providing security assurance within Cloud Computing, has created a Big Data Working Group that has focused on the major challenges to implement secure Big Data services (Cloud Security Alliance, 2013). CSA has categorized the different security and privacy challenges into four different aspects of the Big Data ecosystem. These aspects are Infrastructure Security, Data Privacy, Data Management and, Integrity and Reactive Security. Each of these aspects faces the following security challenges, according to CSA:

• Infrastructure Security

1. Secure Distributed Processing of Data

2. Security Best Actions for Non-Relational Data-Bases

• Data Privacy

3. Data Analysis through Data Mining Preserving Data Privacy

4. Cryptographic Solutions for Data Security

5. Granular Access Control

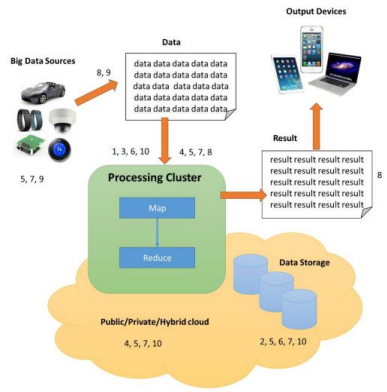
• Data Management and Integrity

6. Secure Data Storage and Transaction Logs

7. Granular Audits 8. Data Provenance

• Reactive Security 9. End-to-End Filtering & Validation

10. Supervising the Security Level in Real-Time These security and privacy challenges cover the entire spectrum of the Big Data lifecycle (Figure 2): sources of data production (devices), the data itself, data processing, data storage, data transport and data usage on different devices



Security and Privacy challenges in Big Data ecosystem (adapted from (Cloud Security Alliance, 2013))

A particular aspect of Big Data security and privacy has to be related with the rise of the Internet of Things (IoT). IoT, defined by Oxford1 as "a proposed development of the Internet in which everyday objects have network connectivity, allowing them to send and receive data", is already a reality – Gartner estimates that 26 billion of IoT devices will be installed by 2020, generating an incremental revenue of $300 billion (Rivera & van der Meulen, 2014). The immense increase in the number of connected devices (cars, lighting systems, refrigerators, telephones, glasses, traffic control systems, health monitoring devices, SCADA systems, TVs, home security systems, home automation systems, and many more) has led to manufacturers to push to the market, in a short period of time, a large set of devices, cloud systems and mobile applications to exploit this opportunity. While it presents tremendous benefits and opportunities for end-users it also is responsible for security challenges. HP recently conducted a study on market-available IoT solutions and concluded that 70% of those contain security problems. These security problems were related with privacy issues, insufficient authorization, lack of transport encryption, insecure web interface and inadequate software protection (HP, 2014). Based on some of these findings, HP has started a project at OWASP (Open Web Application Security Project) that is entitled "OWASP Internet of Things Top Ten" (OWASP, 2014) whose objective is to help IoT suppliers to identify the top ten security IoT device problems and how to avoid them. This project, similar to the OWASP Top 10, identified the following security problems:

- Insecure Web Interface:

which can allow an attacker to exploit an administration web interface (through cross-site scripting, cross-site request forgery and SQL injection) and obtain unauthorized access to control the IoT device.

- Insufficient Authentication/Authorization:

Which can allow an attacker to exploit a bad password policy, break weak passwords and access to privileged modes on the IoT device.

- Insecure Network Services:

which can lead to an attacker exploiting unnecessary or weak services running on the device, or use those services as a jumping point to attack other devices on the IoT network.

- Lack of Transport Encryption:

Allowing an attacker to eavesdrop data in transit between IoT devices and support systems.

- Privacy Concerns:

Raised from the fact the most IoT devices and support systems collect personal data from users and fail to protect that data.

- Insecure Cloud Interface:

without proper security controls an attacker can use multiple attack vectors (insufficient authentication, lack of transport encryption, account enumeration) to access data or controls via the cloud website.

- Insecure Mobile Interface:

without proper security controls an attacker can use multiple attack vectors (insufficient authentication, lack of transport encryption, account enumeration) to access data or controls via the mobile interface.

• Insufficient Security Configurability:

Due to the lack or poor configuration mechanisms an attacker can access data or controls on the device.

• Insecure Software/Firmware:

Attackers can take advantage of unencrypted and unauthenticated connections to hijack IoT devices updates, and perform malicious update that can compromise the device, a network of devices and the data they hold.

• Poor Physical Security:

If the IoT device is physically accessible than an attacker can use USB ports, SD cards or other storage means to access the device OS and potentially any data stored on the device. It is clear that Big Data present interesting opportunities for users and businesses, however these opportunities are countered by enormous challenges in terms of privacy and security (Cloud Security Alliance, 2013). Traditional security mechanisms are insufficient to provide a capable answer to those challenges. In the next section, some of these solutions/proposals are going to be addressed.

It is obvious that organizations desperately require new mechanisms and regulations to guarantee the safety of their systems and data even particularly because traditional techniques are ineffective with respect to Big Data security and privacy challenges. Considering all mentioned here, it is still important to understand that open source or latest technologies might have their own drawbacks such as creating a back door or default credentials; which makes it necessary to carefully consider and make sure that availability, integrity, and confidentiality of data remains intact prior to use of any product

6. BIG DATA SECURITY AND PRIVACY ANALYSIS

There are several techniques used for this purpose (as mentioned throughout this paper) such as encryption, logging, and honeypot detection. Big Data phenomenon is not only faced with security challenges but also data privacy issues. These days many companies are wrestling with privacy challenges and liabilities; however, unlike security, privacy is considered as an asset, which makes it a selling point for both customers and stakeholders . The widespread use of Big Data technologies has resulted in storage and analysis of petabytes of data making information classification even more critical than before. The good news is that Big Data analytics (using more sophisticated pattern analysis and analyzing multiple data) can assist organizations with early stage detection and prevention of advanced threats and malicious intruders Based on the latest news, National Security Agency of the United States (NSA) consistently gathers personal data on people from databases of big companies either active on the internet or in the telecommunication field, violating people's privacy all in the name of protecting US citizens. To deal with such complex challenges, there is a dire need for laws and regulations to enforce clear-cut boundaries in terms of unauthorized access, data sharing and misuse of users; personal data. Based on a study done by the Cloud Security Alliance, security and privacy challenges in Big Data is divided into four categories namely as;

1- Infrastructure Security,

2- Data Privacy,

3- Data Management, and
4- Integrity and Reactive Security as explained below:

1. Infrastructure Security includes distributed programming, nodes, data, internode communication, and security practices for the non-relational data stores.

2. Data Privacy includes privacy preserving data analytics, encryption of data center and access control.

3. Data Management refers to data storage security, logging transactions, the provenance of data and auditing.

4. Integrity and Reactive Security consist of real-time monitoring of data and actions, filtering and validation. Based on all the information mentioned here, it is necessary to

implement authorization and authentication mechanisms for users and applications to control access to sensitive data, also encryption and data masking (anonymization) techniques should be applied to data transfers and datasets [3].

7. CONCLUSION AND FUTURE RESEARCH

The main purpose of Big Data analytics is to gain useful information from a large volume of heterogeneous data. However, having access to large-scale, distributed datasets presents certain privacy and security concerns which we have discussed briefly in this paper. We also investigated how Big Data has different requirements with respect to security and privacy in different areas like data collection, storage, analysis, and transfer. Additionally we have comparatively reviewed a number of studies done on Big Data security and privacy, based on which it was concluded that it is important to consistently monitor network traffic in order to detect suspicious behaviors fast, transferable data must be encrypted with proper standard in accordance with the data type, users and devices need to be granted access to be able to use resources, all communications should take place over secure channels and personal data should be masked prior to the publish of the dataset. Big Data privacy and security is one the most important areas for further discussion and research in the future. It is obvious that now there is a need for the development of new or upgrade of current techniques, technologies, and solutions with respect to the current needs. However as mentioned in the previous section, we need to bear in mind that Big Data can be compared to a loaded gun, it can cause harm if not used in a safe manner with proper regulation, but it can also provide safety and security if it is used correctly. The dramatic increase in the amount of stored and streamed and the ability to analyze it can be utilized greatly in information security areas like detection or prediction of anomalies, intrusion, and fraud simply by examining system, network, and website logs/events/ and traffic. For this purpose, large volume and variety of data associated with network history should be collected, and analyzed for pattern recognition [4]. Some of the advantages of using Big Data includes, System performance without a need to delete cancelled accounts or old logs after a certain period particularly since these might be useful for the purpose of forensic investigations later on, also the ability to run complicated and advanced queries on large and unstructured datasets, real-time decision making ability, automatic defense and risk reduction systems by predicting attacks ahead, and finally faster, better and cheaper security in comparison to traditional methods. Development of proper systems, technologies, and solutions to address challenges associated with big data, can help further mitigate the bottlenecks in the areas of security and privacy, not only for today, but also for future to come.

References:

1.Big Data Working Group, 2013. “Expanded Top Ten Big Data Security and Privacy Challenges”, Cloud Security Alliance, pp. 1-39, Available from: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_ Security_and_Privacy_Challenges.pdf, [Accessed on 3rd August 2016]. [4] Cardenas, A., Manadhata, P. &Rajan, S., 2013.

2. “Big Data Analytics for Security”, IEEE Security & Privacy, Vol. 11, No. 6, pp. 74-76, Available from: IEEE Computer Society Digital Library, [Accessed on 8th August 2016].

3.Inukollu, V., Arsi, S. &Ravuri, S., 2014. “SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING”, International Journal of Network Security & Its Applications (IJNSA), Vol. 6, No. 3, Available from: http://airccse.org/journal/nsa/6314nsa04.pdf, [Accessed on 3rd August 2016]. [16] Ishwarappa, &Anuradha J., 2015.

4.“A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology”, International Conference on Computer Communication and Convergence, Volume 48, pp. 319-324. Available from: http:// www.sciencedirect.com/science/article/pii/S187705091500 6973, [Accessed on 2nd August 2016].