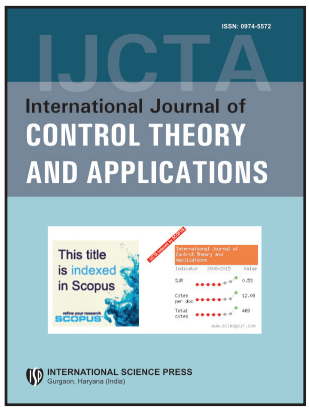

International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 9 • Number 43 • 2016

Security and Privacy Challenges in Big Data Era

Babak Bashari Rad[a], Nafisseh Akbarzadeh[b], Pouya Ataei[c] and Yasaman Khakbiz[d]

[a,c,d]School of Computing, Asia Pacific University of Technology and Innovation (APU), Technology Park Malaysia, Bukit Jalil, Kuala Lumpur, Malaysia. Email: [a]babak.basharirad@apu.edu.my; [c]pouya.ataei.7@gmail.com; [d]yaskhakbiz@gmail.com
[b]School of Technology, Asia Pacific University of Technology and Innovation (APU), Technology Park Malaysia, Bukit Jalil, Kuala Lumpur, Malaysia. Email: nafiseh.akbarzadeh@gmail.com

Abstract: With the explosion of data in world, mainly brought force by the wide spread of Mobile Technologies, Internet of Things (IOT), and Wearable Devices, data analysts are desperately looking for solutions such as Big Data systems to gather more information, discover new opportunities and values, provide in-depth insights, and ultimately increase organizations' bottom line. Big Data phenomenon has touched all aspects of our lives from our Social Media experience to sensitive Medical Records; however, mining and analysing personal information undermines security and privacy causing reputation degradation for the companies using Big Data solutions. This paper explains the definition and characteristics of Big Data, its environment, and highlights latest challenges regarding security and privacy mentioned in the literature such as framework, infrastructure, application, and data. The paper also provides an overall outlook of security and privacy problems in Big Data to make it possible for everyone in the society to take advantage of analysing large datasets without giving up their right to privacy.

Keyword: Big Data, 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 2020 making data production 44 times greater in the year 2020 in comparison to 2009 [7].

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 [9]; 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 [10].

The topic of Big Data has become very important in different areas like science, government, and enterprise to a point that US government has released a series of new reports addressing benefits and some of issues (particularly in relation to security and privacy) resulting from this growth. All the available reports concur

that we need to primarily focus on developing new policies and legal frameworks to foster innovation, promote exchange of information while limiting harm caused by breach of privacy and security to individuals and society [8]. In addition to the development of new policies, wide collection and storage of data has made it necessary to develop scalable tools and technologies like Map Reduce (by Google) and Hadoop (by Apache) to appropriately deal with new data dimensions [5].

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 [9]. 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 [3]. 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 [13].

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 [3].

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 categorize, and investigates the main security and privacy concerns in relation to Big Data within current literature. In Section IV, 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 [25] 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) [13, 29]. Figure 1 shows the different characters of Big Data via 5Vs [24].

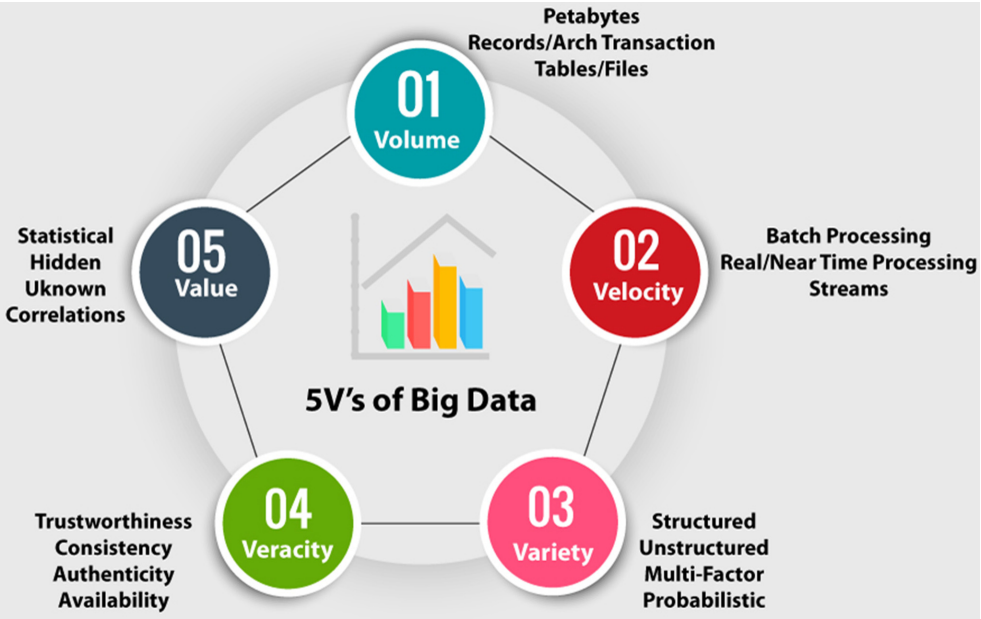

Figure 1: 5Vs of Big Data [24]

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 Wal-Mart, 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 [16].

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 [27]. 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 [13, 27, 29].

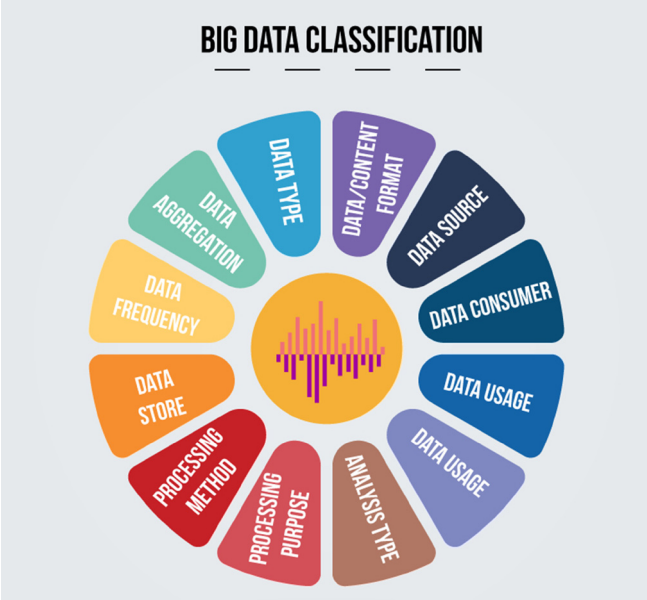

Figure 2: Big Data Classification [13, 27, 29]

3. BIG DATA SECURITY AND PRIVACY

Traditional security and privacy approaches are incapable of fully addressing changes that Big Data has introduced to the digital world, ranging from the amount of data that is collected, stored to its manipulation [42]. Security measures such as complex encryption algorithms, access control limitations, firewalls, and intrusion detection systems for network security can be broken, and even anonymized data could be re-identified and associated with a specific user for malicious use [25].

There are a number new regulations proposed specifically for addressing challenges Big Data has introduced to the privacy of individuals, challenges like, inference and aggregation which makes it possible to re-identify individuals even after identifiers are removed from a dataset; however there are cases in which previously defined regulations may result in privacy violation such as retention of email data for a certain period (in cases up to 5 years) which simply leaves the door open for potential privacy violations [42].

However here we face an old dilemma namely as security triangle; which states that as we employ harder security measures, we negatively affect systems' functionality and ease of use, for example, if a certain regulation limits corporations' access to analysis and manipulation of raw data, corporations would not be able to enhance their business; therefore we are required to propose a balanced approach towards regulations and analytics that ensures corporations' right to analytics as well as individuals' privacy. In a nutshell, the entire ecosystem of Big Data from infrastructure and management to trust policies, integrity, and data quality must be re-visited and further examined in relation to security and privacy concerns [42].

This section we have listed some of Big Data security and privacy issues; however, there is still a need for a comprehensive research to thoroughly identify, and address these concerns. Also, to make sure that security measures are incorporated into all technologies developed for Big Data, such as technologies for infrastructure, monitoring and auditing processes, applications, and data provenance. Here we looked at Big Data (security & privacy) challenges from 5 different perspectives namely as framework (Hadoop), infrastructure (Cloud), monitoring and auditing, key management and data security (anonymization) as it can be seen in Figure 3 [34].

A. Hadoop Security

Hadoop is an open source distributed process framework which utilizes MapReduce model to process large datasets and is widely used by big companies like Google, LinkedIn, Facebook, and Yahoo for data processing [38]; however, this framework was not originally developed for operation in an untrusted environment; therefore, the necessary security measures were not incorporated [1]. Lack of proper security protection in many technologies developed for Big Data such as Hadoop, Twitter Storm, Pig, Hive, MapReduce, Mahout, and Cassandra has turned infrastructure to a security challenge for Big Data management and analytics [42].

However once against all its shortcomings in security, Hadoop received a great degree of interest and was selected as one of the major platform for Big Data, making it compulsory to Figure out the ways in which necessary security precautions can be added especially since hackers usually target data stored on the cloud [1]. Here we mention the two major security weaknesses of Hadoop (1-Acessing Data on Cloud, 2-HDFS Security) and briefly discuss the techniques that can be used while developing a Hadoop system to guarantee data security and privacy:

1. *Hadoop Security and Privacy:* One way in which secure access of users to the data stored on the cloud is provided, is through user authentication prior to granting access to a name node, in this mechanism both user and name node generate a hash function using algorithms such as SHA-256, name node performs a comparison between the hash value sent by the user and the one generated and grants access if the values are correct. This *Trust Mechanism (User & Name Node)* manages to grant access to data nodes [32].

Another easy and commonly used way to ensure data safety and limit unauthorized access is performing encryption and decryption using *Random Encryption Algorithms* like AES, Triple DES, RSA, RC6, IDEA and Rijndael as used by MapReduce [32].

2. *HDFS Security:* This is Hadoop's distributed file system which has three main components namely as name node (master node), data node and secondary name node. HDFS creates several replicas of each block of data in order ensure availability and fast response time. However, HDFS has certain issues with respect to authentication for which use of *Kerberos* (authentication protocol) has been suggested to allow nodes prove their identity to one another [2].

   Another issue that HDFS faces is with regards to naming node's (master name node) unavailability for which the use of an extra name node (slave name node) is suggested that can be accessed in case anything happens to the master name node. The access to the slave node is granted by the administrator if the condition mentioned on *Name Node Security Enhance (NNSE)* holds [12].

To ensure the security of the replicated data and make sure that the access is only granted to the authorized users, *Bullseye* Algorithm is used for data monitoring [35].

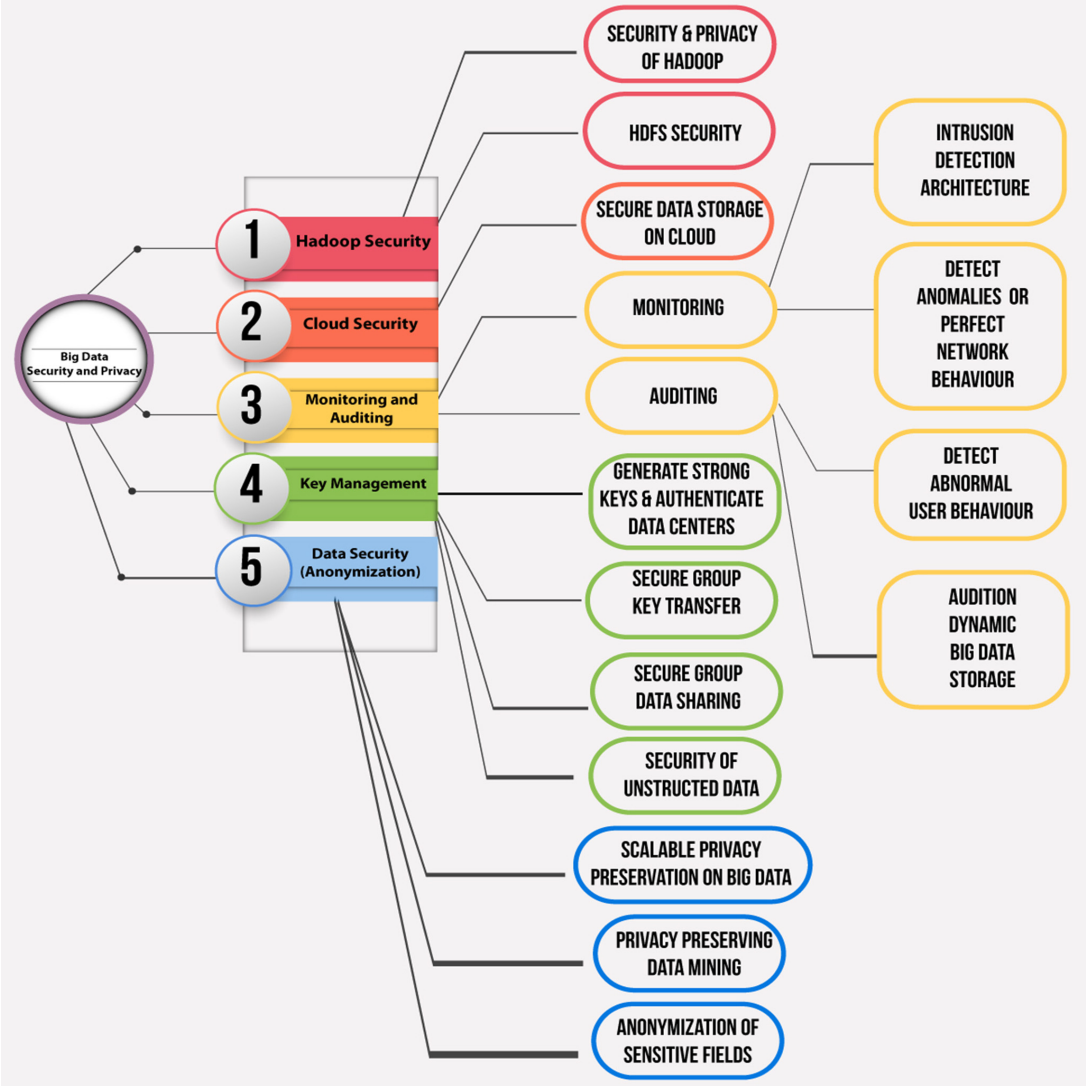

Figure 3: Big Data Security and Privacy Areas [34]

B. Cloud Security

Cloud computing is widely used in association with Big Data due to the numerous advantages it provides namely as on-demand/real-time service availability, widespread access, and sharing of resources [13, 27, 29].

However, usage of cloud computing comes with a huge number of security challenges since this technology includes multiple areas and principals like networking, resource sharing, databases, virtualization, operating systems etc., therefore security issues of these systems and technologies are applicable to cloud computing [33].

One of the main issues with the cloud is securing storage data. Henceforth, cloud service providers have suggested secure ways for sharing Big Data on the cloud platform [19]. These providers assure that their clients do not face issues like data loss or theft, caused by user impersonation [18]. Here we have divided the challenges cloud security into three categories namely as network level, user authentication level, and data level issues.

1. *Network level:* Network level protocol and security issues usually include areas like internode communications and distributed nodes and data [33]; therefore, it is advised to encrypt all network communications by *Secure Sockets Layer (SSL)* to guarantee the security of packets and assure that No. useful information can be derived even if an unauthorized user gains access to the network communications [15].

2. *Authentication level:* User authentication level security issues usually include areas like authentication methods like logging, nodes' administrative permissions, applications' authentication and techniques used for encryption/decryption. To address these problems, it is important to always *Log Data Modification* activities performed by the users and regularly audit them to see if data was manipulated. In addition to that, it is important to *Validate Nodes' Authenticity* using technologies like Kerberos prior to joining a cluster and, as a secondary measure, set some *Honey Pot Nodes* within clusters to trap hackers in case they were successful to pass the authentication [33].

3. *Data level:* Data level security issues usually include areas like distributed data protection to assure data availability and integrity [33]. It is important to always have a minimum of three different *Back-Up Servers*, ready to go online, in case the main server becomes unavailable due to technical problems, attacks, or natural disasters. In addition to that distributed data should always be kept in the encrypted and compressed form to avoid security issues [6].

The *Cryptography* scheme used here takes advantage of virtual mapping to divide the data into different parts and place it on multiple storages to make it impossible for hackers to gain complete access to it [34].

C. Monitoring and Auditing

Monitoring and auditing is an integral part of network security management; which helps the service providers to prevent security breaches simply by checking network traffic and using the information acquired to adjust or apply certain security measures. While network monitoring mainly focuses on collecting and studying events to foresee/detect intrusions; network auditing is considered as a systematic and measurable security policy, which has a great impact on network security [43].

1. *Network Monitoring:* There are factors that need to be analysed if an *Intrusion Detection and Prevention Architecture* is to be applied successfully monitoring the entire network traffic, some of the areas that must be included in the monitoring process include HTTP and DNS traffic, records of the IP flow, and data gathered by the honeypots laid as a trap for intruders. The suggested Intrusion Detection System (IDS) stores and processes data using three *Malicious Likelihood Metrics* to find

out whether packets, flow or domain name is affected in which case the action/process will be stopped immediately [23].

Detect Abnormal User Behaviour and suspicious data behaviour is another important area within the network for which a *Self-Assuring System* has been presented. This system keeps a library of keywords associated with untrusted behaviour, and generates a low critical log of users' identification information performing a suspicious action based on which a second log is produced (high critical log) by checking the occurrence frequency of the low critical logs and deciding whether it has reached it maximum limitation. In the last step, the self-assuring system blocks suspicious user's access to the network [11].

Due to the heterogeneous nature of Big Data, it is crucial to deploy a monitoring system capable of *Detecting Anomalies* from within the data stream, simply by collection of the network logs, classification/filtering of them, and finally analysing and acquiring meaning/correlation of this data based on which necessary statistics are generated and correct *Predictions* are formed in relation to the *Network Behaviour* and events [20].

2. *Network Auditing:* We all know that Big Data (with its specific characteristics) has deeply impacted the way data analytics is done in general; however, there are some challenges in respect with data auditing when it comes to Big Data integrity and availability.

   While Big Data availability can be easily achieved by keeping a number of replicas to assure easy and fast access, it might cause some issues with data integrity especially since overhead for update verification of dynamic datasets is huge and there is No. integrity scheme for simultaneous auditing and authentication of block indices; Therefore here we suggest use of *MuR-DPA* which is a top-down, multi-replica public auditing scheme based on *Merkle Hash* Tree incorporating an authenticated data structure to conduct a secure public auditing for *Dynamic Big Data Storage* on cloud [21].

D. Key Management

Security and privacy enhancement of Big Data come with a different number of challenges especially since dynamic key generation for the Big Data is not efficient using current cryptographic techniques. In today's mobile society, data centre users can be located anywhere, which makes it necessary to have a specific key management system to simultaneously secure the data and channel used for transmission between nodes. To solve this problem, data centres need efficient *Quantum Cryptography* using Grover's algorithm for appropriate authentication approaches to enhance the security and privacy with less complexity in mobile or fixed data centres [41].

Even though use of Quantum model can increase system efficiency (by lowering the number of key search operations) and security (decline in number of attacks); it is critical to remember that Big Data communications require *Secure Group Key Transfer* protocols to withstand attacks; therefore, using an online key generation centre based on *Diffie-Hellman Key Agreement* is suggested [14].

Another aspect that we should look at here is *Secure Group Data Sharing* for which Conditional Proxy Re-Encryption (CPRE) is used enabling group sharing of sensitive data without exposing the actual content decryption key to people who are not within the share group. Here, we introduce the latest scheme for secure group data sharing on the cloud which is more compatible and efficient with Big Data called the *Outsourcing CPRE scheme (O-CPRE)* which decreases the client overhead greatly [39].

Big Data can be broadly divided into two general type namely as structured and unstructured, and as you might already guess it is far more difficult to assure *Security for Unstructured Data*. Here we would suggest on what can be used to guarantee the security of unstructured data.

In this approach, data is reviewed, filtered, clustered, and finally classified based on its type and level of sensitivity, afterward specific data nodes are created in the database. To provide security to data nodes, a security suite was designed which incorporates different security standards and algorithms in accordance with the type of data node. At this stage, the most suitable algorithm is assigned to the data node based on the type of data/its requirements (confidentiality, authentication, and integrity) and sensitivity level (sensitive, confidential, public) from security suite [17].

E. Data Security (Anonymization)

Data collection and harvesting for analysis has raised too many eyebrows with respect to the user right to the privacy. One of the most important responsibilities of data publishers is to assure data security and privacy; even though this might prove to be unachievable at certain times. Privacy Preserving Data Publishing (PPDP) discusses the ways in which data can be published while ensuring users' right to privacy [37]. Even though it is becoming more and more difficult for the data publishers to mask Personally Identifiable Information (PII) due to the speed in which the data is shared, there is a dire need to devise policies in which companies are held accountable to ensure the anonymization (de-identified) and secure transfer of users' personal data [40].

Unfortunately, even after performing the anonymization process, there are ways (use of strong algorithms and the artificial intelligence analysis) in which users can be re-identified. To avoid this issue, *K-Anonymity based Metrics* were used to mask sensitive fields. Here personal identifiers are removed from usage logs to protect users' privacy. The anonymization of sensitive fields is achieved by use of AES symmetric key encryption, which is stored in HDFS for analysis. In situations in which re-identification of data is needed, stored logs are moved and decrypted using the key [36].

Privacy Preserving Data Mining (PPDM) is another subject that has gained traction due to increased use of analytics and privacy concerns. It is crucial to gain privacy without compromising data content or mining accuracy; therefore, here we advise the use of an algorithm named *Adaptive Utility based Anonymization (AUA)* model to address the risk of data disclosure without affecting classification accuracy [28].

Scalability and huge volume is another reason that normal anonymization methods are unsuccessful in masking sensitive information when it comes to Big Data; therefore, here we advise use of a *Hybrid Top-Down & Bottom-Up Subtree Anonymization* model to increase scalability capability of the method [45].

However, based on a different study, there is a different option to increase scalability called a *Two-Phase Clustering Algorithm*. This approach includes a t-ancestors clustering algorithm and a proximity-aware agglomerative clustering algorithm that has been designed with MapReduce to gain higher scalability. This approach improves scalability and the time-efficiency of local-recoding anonymization in addition to the ability to defend against proximity privacy breaches [44].

4. BIG DATA SECURITY AND PRIVACY ANALYSIS

As mentioned in this paper, a great number of the businesses utilize Big Data for marketing and research purpose; however, most of them lack fundamental assets when it comes to security; which might lead to serious lawsuits and reputational damage if a security breach occurs [15]. Therefore, 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

[26]. There are several techniques used for this purpose (as mentioned throughout this paper) such as encryption, logging, and honeypot detection [18].

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 [15]. 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 [15].

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 [25]. 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].

5. CONCLUSION AND FUTURE RESEARCH

The main purpose of Big Data analytics is to gain useful information from a large volume of heterogeneous data [30]. 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 [4, 20, 22, 31]. 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] Adluru, P., Datla, S. & Zhang, X. 2015. "Hadoop eco system for big data security and privacy", 2015 Long Island Systems, Applications and Technology, pp. 1-6, Available from: IEEE Computer Society Digital Library, [Accessed on 1st August 2016].

[2] Alam, A. & Ahmed, J., 2014. "Hadoop architecture and its issues", Proceedings - 2014 International Conference on Computational Science and Computational Intelligence, CSCI 2014, Vol. 2, pp. 288-291, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].

[3] 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. "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].

[5] Chen, P. & Zhang, C., 2014. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Sciences, Vol. 275, pp. 314-347, Available from: Science Direct, [Accessed on 5th August 2016].

[6] Cheng, H. & Rong, C. & Hwang, K. & et. al., 2015. "Secure big data storage and sharing scheme for cloud tenants", China Communications, Vol. 12, No. 6, pp. 106-115, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].

[7] CSC, 2012. "The rapid growth of global data", Available from: https://assets1.csc.com/insights/downloads/CSC_Infographic_Big_Data.pdf, [Accessed on 1st August 2016].

[8] Gaff, B., Egan, H. & Geetter, J., 2014. "Privacy and Big Data", pp. 7-9, Available from: IEEE Computer Society Digital Library, [Accessed on 2nd August 2016].

[9] Gheid, Z. & Challal, Y., 2015. "An Efficient and Privacy-Preserving Similarity Evaluation for Big Data Analytics", 2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC), No. 1, pp. 281-289, Available from: IEEE Computer Society Digital Library, [Accessed on 2nd August 2016].

[10] Guerra, E., De Lara, J. & Malizia, A. et. al., 2012. "Supporting user-oriented analysis for multi-view domain-specific visual languages", Information and Software Technology, Vol. 51, No. 4, pp. 769-784, Available from: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf , [Accessed on 1st August 2016].

[11] Gupta, A., Verma, A., Kalra, P. & et. al., 2015. "Big Data: A security compliance model", Proceedings of the 2014 Conference on IT in Business, Industry and Government: An International Conference by CSI on Big Data, Available from: IEEE Computer Society Digital Library, [Accessed on 4th August 2016].

[12] Hadoop, 2016. "HDFS User Guide", Available from: https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsUserGuide.html, [Accessed on 3rd August 2016].

[13] Hashem, I., Yaqoob, I. & Anuar, N. et. al., 2015. "The rise of "big data" on cloud computing: Review and open research issues", Information Systems, Vol. 47, pp. 98-115, Available from: Science Direct, [Accessed on 3rd August 2016].

[14] Hsu, C., Zeng, B. & Zhang, M., 2014. "A novel group key transfer for big data security", Applied Mathematics and Computation, Vol. 249, No. 2013, pp. 436-443, Available from: Science Direct, [Accessed on 7th August 2016].

[15] 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. "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/S1877050915006973, [Accessed on 2nd August 2016].

[17] Islam, R. & Islam, E., 2014. "An approach to provide security to unstructured Big Data", SKIMA 2014 - 8th International Conference on Software, Knowledge, Information Management and Applications, Available from: IEEE Computer Society Digital Library, [Accessed on 7th August 2016].

[18] Jain, P, 2012. "Security issues and their solution in cloud computing", International Journal of Computing and Business Research, Available from: http://www.researchmanuscripts.com/isociety2012/1.pdf, [Accessed on 3rd August 2016].

[19] Kumar, A. & Lee, H., n.d. "Efficient and Secure Cloud Storage for Handling Big Data", No. 3, pp. 162-166, Available from: IEEE Computer Society Digital Library, [Accessed on 1st August 2016].

[20] Lan, L. & Jun, L., 2013. "Some Special Issues of Network Security Monitoring on Big Data Environments", 2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing, No. 2012, pp. 10-15, Available from: IEEE Computer Society Digital Library, [Accessed on 4th August 2016].

[21] Liu, C., Ranjan, R., Yang, C. & et. al., 2015. "MuR-DPA: Top-Down Levelled Multi-Replica Merkle Hash Tree Based Secure Public Auditing for Dynamic Big Data Storage on Cloud", IEEE Transactions on Computers, Vol. 64, No. 9, pp. 2609-2622, Available from: IEEE Computer Society Digital Library, [Accessed on 4th August 2016].

[22] Mahmood, T. & Afzal, U., 2013. "Security Analytics: Big Data Analytics for Cybersecurity", 2013 2nd National Conference on Information Assurance (NCIA), pp. 129-134, Available from: IEEE Computer Society Digital Library, [Accessed on 8th August 2016].

[23] Marchal, S., Jiang, X., State, R. & et. al., 2014. "A big data architecture for large scale security monitoring", Proceedings – 2014 IEEE International Congress on Big Data, pp. 56-63, Available from: IEEE Computer Society Digital Library, [Accessed on 4th August 2016].

[24] Marr, B., 2015. "Why only one of the 5 Vs of big data really matters", Available from: http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters, [Accessed on 4th August 2016].

[25] Matturdi, B., Zhou, X., Li, S. & et. al., 2014. "Big Data security and privacy: A review", China Communications, Vol. 11, No. 14, pp. 135-145, Available from: IEEE Computer Society Digital Library, [Accessed on 7th August 2016].

[26] Miloslavskaya, N., Senatorov, M., Tolstoy, A. & et. al., 2014. "Information security maintenance issues for big security-related data", Proceedings - 2014 International Conference on Future Internet of Things and Cloud, FiCloud 2014, pp. 361-366, Available from: IEEE Computer Society Digital Library, [Accessed on 7th August 2016].

[27] Mysore, D. & Khupat, S. & Jain, S., 2013. "Introduction to Big Data classification and architecture". Available from: https://www.ibm.com/developerworks/library/bd-archpatterns1/, [Accessed on 1st August 2016].

[28] Panackal, J. & Pillai, A., 2015. "Adaptive Utility-based Anonymization Model: Performance Evaluation on Big Data Sets", Procedia Computer Science, Vo. 50, No. 2, pp. 347-352, Available from: Science Direct, [Accessed on 6th August 2016].

[29] Perreault, L., 2015. "Big Data and Privacy: Emerging Issues", Conf-IRM 2015 Proceedings, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].

[30] Philip Chen, C. & Zhang, C., 2014. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Information Sciences, Vol. 275, pp. 314-347, Available from: Science Direct, [Accessed on 8th August 2016].

[31] Pratyusa K., Alvaro A. & Sreeranga, P., 2013. "Big Data Analytics for Security Intelligence", IEEE Security & Privacy, pp. 74-76, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].

[32] Saraladevi, B., Pazhaniraja, N. & Paul, P. et. al., 2015. "Big data and Hadoop-A study in security perspective", Procedia Computer Science, Vol. 50, pp. 596-601, Available from: Science Direct, [Accessed on 4th August 2016].

[33] Saranya, R. & MuthuKumar, V.P., 2015. "Security issues associated with big data in cloud computing", International Journal of Multidisciplinary Research and Development, Vol. 2, No. 4, pp. 580-585, Available from: www.allsubjectjournal.com/archives/download?id=716&refnum=253, [Accessed on 5th August 2016].

[34] Sayed, E., Ahmed, A. & Saeed, R., 2014. "A Survey of Big Data Cloud Computing Security", International Journal of Computer Science and Software Engineering (IJCSSE), Vol. 3, No. 1, pp. 78-85, Available from: http://ijcsse.org/published/volume3/issue1/p3-V3I1.pdf, [Accessed on 3rd August 2016].

[35] Seagle, R., 2012. "A Framework for File Format Fuzzing with Genetic Algorithms", Available from: http://trace.tennessee.edu/cgi/viewcontent.cgi?article=2402&context=utk_graddiss, [Accessed on 3rd August 2016].

[36] Sedayao, J., Bhardwaj, R. & Gorade, N., 2014. "Making big data, privacy, and anonymization work together in the enterprise: Experiences and issues", Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014, pp. 601-607, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].

[37] Singh, A. & Parihar, D., 2013. "A Review of Privacy Preserving Data Publishing Technique", International Journal of Emerging Research in Management &Technology, Vol. 9359, No. 6, pp. 32-38, Available from: http://www.ermt.net/docs/papers/Volume_2/issue_6_June2013/V2N6-132.pdf, [Accessed on 7th August 2016].

[38] Shi, X., Chen, M., He, L. & et. al., 2015. "Mammoth: Gearing Hadoop towards Memory-Intensive MapReduce Applications", IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 8, pp. 2300-2315, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].

[39] Son, J., Kim, D., Hussain, R. & et. al., 2014. "Conditional proxy re-encryption for secure big data group sharing in cloud environment", Proceedings - IEEE INFOCOM, pp. 541-546, Available from: IEEE Computer Society Digital Library, [Accessed on 7th August 2016].

[40] Tene, O. & Polonetsky, J., 2013. "Big data for all: Privacy and user control in the age of analytics", Northwestern Journal of Technology and Intellectual Property, Vol. 11, No. 5, pp. 240-273, Available from: http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip, [Accessed on 6th August 2016].

[41] Thayananthan, V. & Albeshri, A., 2015. "Big data security issues based on quantum cryptography and privacy with authentication for mobile data center", Procedia Computer Science, Vol. 50, pp. 149-156, Available from: Science Direct, [Accessed on 7th August 2016].

[42] Thuraisingham, B., 2015. "Big Data - Security and Privacy", Proceedings of the 5th ACM Conference on Data and Application Security and Privacy - CODASPY '15, pp. 279-280, Available from: IEEE Computer Society Digital Library, [Accessed on 1st August 2016].

[43] Tsunoda, H., Keeni, G., 2012. "Security by Simple Network Traffic Monitoring", Proceedings of the Fifth International Conference on Security of Information and Networks, Available from: ACM Digital Library, [Accessed on 4th August 2016].

[44] Zhang, X., Dou, W., Pei, J. & et. al., 2015. "Scalable Proximity-Aware Local-Recoding Anonymization for Big Data Privacy Preservation with MapReduce in Cloud", Vol. 64, No. 8, pp. 1-14, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].

[45] Zhang, X., Liu, C., Nepal, S. & et. al., 2013. "Combining top-down and bottom-up: Scalable sub-tree anonymization over big data using mapreduce on cloud", Proceedings - 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2013, pp. 501-508, Available from: IEEE Computer Society Digital Library, [Accessed on 6th August 2016].