A Survey on Security and Privacy Issues in Big Data

Duygu Sinanc Terzi
Gazi University
Computer Engineering
Ankara, Turkey
duygusinanc@gazi.edu.tr

Ramazan Terzi
Gazi University
Computer Engineering
Ankara, Turkey
ramazanterzi@gazi.edu.tr

Seref Sagiroglu
Gazi University
Computer Engineering
Ankara, Turkey
ss@gazi.edu.tr

Abstract—**Due to the reasons such as the rapid growth and spread of network services, mobile devices, and online users on the Internet leading to a remarkable increase in the amount of data. Almost every industry is trying to cope with this huge data. Big data phenomenon has begun to gain importance. However, it is not only very difficult to store big data and analyse them with traditional applications, but also it has challenging privacy and security problems. For this reason, this paper discusses the big data, its ecosystem, concerns on big data and presents comparative view of big data privacy and security approaches in literature in terms of infrastructure, application, and data. By grouping these applications an overall perspective of security and privacy issues in big data is suggested.**

Keywords—big data; Hadoop security; cloud security; monitoring; auditing; key management; anonymization

I. Introduction

Data generation and collection quickly surpass the bounds in the digital universe of today. The data has been doubling every 2 years since 2011 [1]. It is predicted that the data will increase 300 times, from 130 exabytes in 2005 to 40,000 exabytes in 2020 [2]. As a result of this technological revolution, the big data is becoming increasingly an important issue in the sciences, governments, and enterprises. Big Data is a data set, which is difficult to capture, store, filter, share, analyse and visualize on it with current technologies [3].

Despite such difficulties, if you can cope with big data, it provides you with generating revenue, executive efficiency, strategic decisions, better services, defining needs, identifying new trends, and developing new products, all of which is covered in the data science [3]. In addition, data science studies parallel and distributed processing, similarity search, graph analysis, clustering, stream processing, search ranking, association analysis, dimensionality reduction and machine learning algorithms [4]. However, in this complex computation environment, traditional security and privacy mechanisms are insufficient to analyse big data. This challenges in big data consist of computation in distributed and non-relational environments, cryptography algorithms, data provenance, validation and filtering, secure data storage, granular access control, and real time monitoring [5].

Identifying the sources of problems will result in more efficient use of big data. For this reason, this paper examines and classifies studies on security and privacy breaches and solutions in big data. This perspective would lead to an understanding of important research areas and the development of new methods. In addition, the use of big data in analysis would make the systems become safer.

Section II presents a brief summary of big data. Section III contains categorization of big data concerning security and privacy studies in literature. The results obtained with security and privacy issues in big data are discussed in Section IV, and section V explains how to use big data to maintain security. Finally the conclusion highlights the importance and requirements to secure big data communication.

II. Definition and Characteristics of Big Data

Big data refers to large and complex datasets that typical software is inadequate for managing [2]. There are various explanations of big data via Vs. 5Vs are typically used to characterize of Big Data as volume, velocity, variety, veracity and value (Fig. 1) [3,6,7]. Volume is the size of data; velocity is the high speed of data; variety indicates heterogeneous data types and sources; veracity describes consistency and trustworthy of data; and value provides outputs for gains from large data sets.

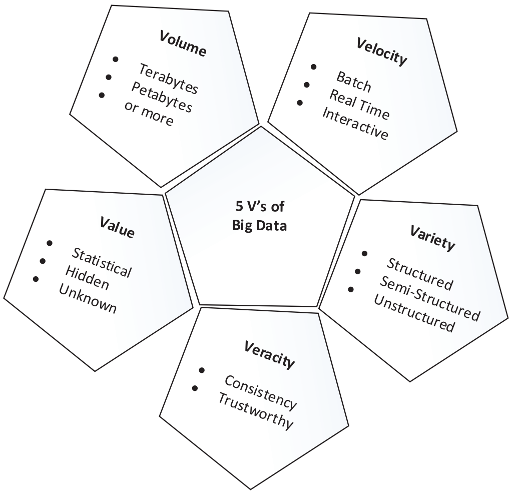

Figure 1. 5 V's of big data

Identifying characteristics of the data is helpful in extracting its hidden patterns. Big data is classified into ten categories in terms of data type, data format, data source, data consumer, data usage, data analysis, data store, data frequency, data processing propose, and data processing method (Fig. 2) [6,8].

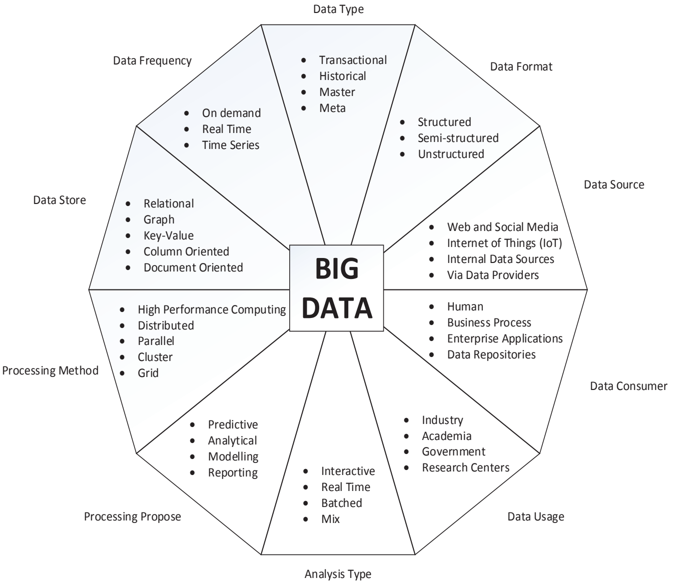

Figure 2. Big data classification

III. Big Data Security and Privacy Approaches in Literature

Traditional solutions are insufficient when dealing with big data to ensure security and privacy. Encryption schemes, access permissions, firewalls, transport layer security can be broken; provenance of data can be unknown; even anonymized data can be re-identified [2]. For these reasons, advanced techniques and technologies are developed to protect, monitor and audit big data processes in terms of infrastructure, application and data. Considering the related literature, this paper has categorized security and privacy issues for big data under 5 titles as Hadoop security, cloud security, monitoring and auditing, key management and anonymization (Fig. 3). Table I summarizes the studies in terms of purpose, method, and data.

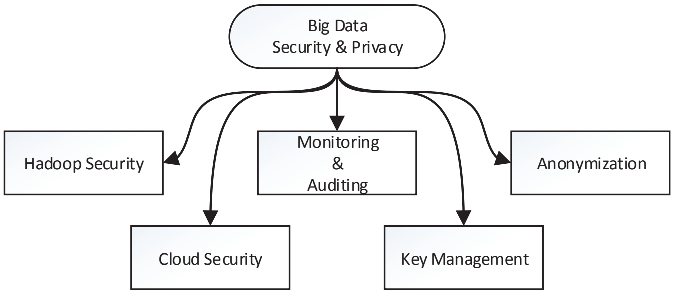

Figure 3. Big data security and pravicy categorization

A. Hadoop Security

Hadoop is a distributed process framework and it was not originally developed for security. It was meant to operate in trusted environments. As Hadoop has become a popular platform, security precautions have started to be developed. In addition, it has started to receive academic interest.

When developing a Hadoop system that guarantees security and privacy of information on the cloud, two techniques were proposed to prevent a hacker who wants to get all data in cloud [7]. A trust mechanism has been implemented between user and name node which is component of HDFS and manages data nodes. According to this mechanism user must authenticate himself to access name node. Firstly, user sends hash function then name node produces hash function too and it compares these two generated functions. If compare result is correct, accessing system is provided. In this step, SHA-256 which is one of the hashing techniques is used for authentication. Random encryption techniques such as RSA, Rijndael, AES and RC6 has been also used on data in order that a hacker does not gain an access whole data. MapReduce is executed encryption/decryption process in this approach. Finally, these two techniques are tested using twitter stream for indicating how to maintain of security issues.

Another unit that cause the security weakness is Hadoop Distributed File System (HDFS). Three methods to increase HDFS security has been developed [9]. In order to achieve authentication issue, Kerberos mechanism based on Ticket Granting Ticket or Service Ticket have been used as first method. The second method is about monitoring all sensitive information in 360° by using Bull Eye algorithm. This algorithm has been used to make sure data security and manage relations between original data and replicated data. It is also allowed only authorized person to read or write critical data. To handle name node problems as final method, two name node has been proposed: one of them is master and the other is slave. If something happened to master node, administrator gives data from slave name node on condition that Name Node Security Enhance (NNSE) permission. Therefore latency and data availability problems succeeded in secure way.

B. Cloud Security

The widespread use of cloud computing for such reasons as broad network access, on-demand service, resource pooling and being elastic have made a proper environment for big data [6]. However, cloud hosts traditional threats and new attacks.

Data storage on clouds is one of the main problems nowadays. Therefore, some precautions must be taken by the service provider. Because of this, a secure way to handle and share big data on cloud platform has been presented [10]. It includes many security methods like authentication, encryption, decryption, and compression etc. to store big data securely. Authentication with email and password has been used for the authorized person. Data has been encrypted and compressed to prevent security issues. It also takes precautions in case of a natural disaster and uses three backup servers for this purpose. In these servers, data has been stored in an

TABLE I. *CATEGORIZATION OF BIG DATA SECURITY AND PRAVICY STUDIES*

Studies	Purpose		Method	Data
[7]	Hadoop Security	Security and privacy of Hadoop	Trust mechanism between user and name node, random encryption technique on data	Twitter stream
[9]		HDFS Security	Kerberos mechanism, Bull Eye Algorithm and Name node approach	N/A*
[10]	Cloud Security	Secure data storage on cloud	Authentication, encryption, decryption, and compression	N/A
[11]		Secure data storage on cloud	Key establishment scheme, and identity based encryption algorithm	10GB sized dataset
[12]	Monitoring	Intrusion detection architecture	Maliciousness likelihood metrics	DNS data, IP flow records, HTTP traffic, honeypot data
[13]		Detect anomalies or predict network behaviour	Data collection, integration, analysis, and interpretation	N/A
[14]		Detect abnormal user behaviour	Self-assuring system	N/A
[15]	Auditing	Auditing Dynamic Big Data Storage	MuR-DPA based on Merkle Hash Tree	N/A
[1]	Key Management	Generate strong keys, and authenticate data centres	Quantum cryptography and PairHand protocol	N/A
[16]		Secure group key transfer	Online key generation centre based on Diffie-Hellman key agreement and linear secret sharing scheme	N/A
[17]		Secure group data sharing	Outsourcing conditional proxy re-encryption scheme	N/A
[18]		Security of unstructured data	Data analytics (data filtering, clustering, classification) and security suit (existing security standards and algorithms)	XML, e-mail, text, image, video, audio etc,
[20]	Anonymization	Anonymization of sensitive fields	K-anonymity based metrics	Intel Circuit logs
[21]		Privacy preserving data mining	Adaptive Utility-based Anonymization Model	UCI Adult Data set
[22]		Increasing scalability capability for anonymization	Hybrid Top-Down and Bottom-Up Sub-Tree Anonymization	Adult data set
[23]		Scalable privacy preservation on big data	Two-phase clustering algorithm	Cencus-Income

*N/A: Not Applicable

encrypted format. If something happens to the server, encrypted data has been decrypted with the secret key.

The classical encrypted technique is not enough for big data security on cloud. Consequently, new scheme to secure big data storage has been proposed [11]. This scheme uses cryptographic virtual mapping to create data path. According to the proposed scheme, big data has been separated into many parts and each part is located in different storage providers. As a security measure, if all data encryption are thought to be quite computational and useless, only storage path which shows critical information encryption seems enough, rather than all big data encrypts. The proposed scheme also supports some information encryption to increase the security level. To achieve availability, the scheme holds multiple copies of each part and their accessing index. Thus, if any data part is lost for some reason, information availability is successfully maintained.

C. Monitoring and Auditing

Security monitoring is gathering and investigating network events to catch the intrusions. Security audit is a systematic measurable security policy to use different methods. These two elements play an important role in active security.

Intrusion detection and prevention procedures on the whole network traffic is quite difficult. To solve this problem, a security monitoring architecture has been developed via analysing DNS traffic, IP flow records, HTTP traffic and honeypot data [12]. The proposed solution includes storing and processing data in distributed sources through data correlation schemes. At this stage, three likelihood metrics have been calculated to identify whether domain name, packet or flow is malicious. According to the score obtained through this calculation, an alert occurs in detection system or process terminate by prevention system. According to performance analysis with open source big data platforms on electronic payment activities of a company data, Spark and Shark produce fast and steady results than Hadoop, Hive and Pig.

Network security systems for big data should be find abnormalities quickly and identify correct alerts from heterogeneous data. Therefore, a big data security event monitoring system model has been proposed which consists of four modules: data collection, integration, analysis, and interpretation [13]. Data collection includes security and network devices logs and event information. Data integration process is performed by data filtering and classifying. In data analysis module, correlations and association rules are determined to catch events. Finally, data interpretation provides visual and statistical outputs to knowledge database that makes decisions, predict network behavior and respond events.

The separation of non-suspicious and suspicious data behaviour is one other issue of monitoring big data. Therefore, a self-assuring system which includes four modules has been suggested [14]. The first module contains keywords that are related to untrusted behaviour and it is called library. The second module records identification information about event when a suspicious behaviour occurs and this step is named as a low critical log. High critical log as the third module counts low critical logs' frequency and checks whether low critical logs reach the thresholds value. The last module is a self-assuring system and the user is prevented by the system if he/she has been detected as suspicious.

While big data becomes a new phenomenon with 5V features, new gaps are emerging for big data auditing such as data availability, consistency, integrity, identification, aggregation and confidentially. Hence, some precautions must be taken for all of these gaps in terms of big data. Data availability is satisfied with multiple replicas on big data environment [15]. Thanks to replica nodes, accessing information is quite easy and fast even though some data nodes may be damaged for any reason. These advantages sound good, but they lead to a few security problems like data integrity trouble. Data integrity is that ensuring data is recorded right roughly. Classical method to make sure data integrity is that getting all data blocks from the server and has been verified by client. However, this way is inapplicable on big data space. Thus, to audit dynamic big data storage, some research has been conducted. In [15], communication overhead and public auditing and authentication problems have been solved with proposed scheme based on Multi-Replica Merkle Hash Tree.

D. Key Management

Key generating and sharing between servers and users is another big data security issue. However, using big data centres, quick and dynamic authentication protocols can be suggested.

In [1], a layered model has been proposed for quantum cryptography for strong keys in less complexity and PairHand protocol for authentication in mobile or fixed data centres. The model consists of these layers: front end, data reading, quantum key processing, quantum key management and application layers, respectively. This model has been not only increased efficiency but also reduced key search operations and passive attacks.

The big data services consist of multiple groups that need group key transfer protocols for secure communications. For this reason, novel protocol without an online key generation centre based on Diffie-Hellman key agreement and linear secret sharing scheme unlike existing protocols has been offered [16]. The protocol counter attacks via ensured key freshness, key authentication and key confidentiality reducing system overhead.

In more complex systems, conditional proxy re-encryption (CPRE) is used for secure group data sharing. Accordingly, an outsourcing CPRE scheme has been proposed in cloud environment which reduces overhead without downloading all data from the cloud, encrypting them and uploading them to the cloud in a new condition unlike CPRE [17]. When a group membership has been changed, key generation and decryption processes execute on outsourcing server and a condition value changing key has been calculated. Then it is sent to the cloud. After that, the cloud storage uses this key to transform existing data.

Due to the variety of big data, ensuring the safety of the unstructured data like text, e-mail, XML or media is more difficult than the structured data. Therefore, a security suit has been developed for data node consisting of different types of data and security services for each data type [18]. The proposed approach contains two stages, data analytics, and security suite. Firstly, filtering, clustering and classification based on data sensitivity level is done in data analytics phase. Then data node of databases is created and a scheduling algorithm selects the appropriate service according to section (identification, confidentiality, integrity, authentication, non-repudiation) and sensitivity level (sensitive, confidential, public) from security suite. For example, to provide privacy of sensitive text data, 3DES algorithm is selected.

E. Anonymization

Data harvesting for analytics causes big privacy concerns. Protecting personally identifiable information (PII) is increasingly difficult because the data are shared too quickly. To eliminate privacy concerns, the agreement between the company and the individual must be determined by policies. Personal data must be anonymized (de-identified) and transferred into secure channels [19]. However, the identity of the person can be uncovered depending on the algorithms and the artificial intelligence analysis of company. The predictions made by this analysis can lead to unethical issues.

In [20], PII has been removed from Intel Circuit web portal usage logs to protect users' privacy. The proposed architecture makes anonymization of sensitive fields in log data with AES symmetric key encryption and stores it in HDFS for analysis. When de-anonymization is needed, the logs are moved back and the masking areas are decrypted with the same key. Lastly, the quality of anonymization is measured by k-anonymity based metrics.

With the increase of individual and organizational privacy concerns, Privacy Preserving Data Mining (PPDM) has begun to gain tremendous importance. However, these techniques affect the success of applications. To provide privacy protection, an Adaptive Utility based Anonymization (AUA) has been proposed, which depends on association mining [21]. Both naïve and masked data sets has been tested. The results show that anonymization has not been a cause of a critical decrease in classification accuracy with this iterative process.

There are many classical methods to fulfil anonymization over data, but none of them is sufficient for big data because they suffer from scalability issues because of the volume of the data [22]. The classical anonymization methods must be rearranged to handle big data anonymization problem. Consequently, a hybrid scheme has been proposed which combines two classical method such as Top-Down and Bottom-up for Sub-Tree Anonymization to raise scalability capabilities on big data using MapReduce. The suggested scheme has been tested and the results show that hybrid sub-tree approach has better performance than classical sub-tree anonymization. In another study, when compared with [22], a new scalable method for local recording scheme considering the proximity-aware privacy has been proposed in [23]. In this scheme, data sets have been generated at cell level. To solve scalability problem, two steps have been planned and coded for MapReduce jobs. The first step is used to split dataset using t-ancestor clustering; the second step records data with the proximity-aware agglomerative algorithm.

IV. Security and Privacy in Big Data

Seeking new ways to take advantage of big data, organizations need secure mechanisms and regulations to guarantee their systems. It is thought that the traditional techniques are ineffective in big data security and privacy issues. Nevertheless open source or new technologies (if they are not well understood) also host unknown back doors and default credentials [4]. Therefore, confidentiality, integrity and availability of information must be carefully considered.

A. Security

Diversity of data sources, data formats, streaming of data and infrastructures may cause unique security vulnerabilities. The Cloud Security Alliance has divided security and privacy challenges in big data into four categories; infrastructure security, data privacy, data management, integrity, and reactive security [5]. Infrastructure security consists of secure distributed programming and security practices in non-relational data stores. Data privacy refers to privacy preserving analytics, encrypted data centre and granular access control. Data management involves secure data storage and transaction logs, auditing and data provenance. In addition, integrity and reactive security include validation, filtering and real time monitoring. On the basis of these proposed issues, authorization and authentication mechanisms must be constituted for both users and applications, and encryption and data masking must be implemented for both data rest and stream.

B. Privacy

The development of systems and applications has led to the termination of the individual control about collection and usage of PII. According to the latest news, National Security Agency (NSA) eavesdropped personal data from heterogeneous data sources such as databases of big companies, internet and telecommunication under cover of protecting US citizens [2]. Many big data projects like this indicate the violation of people's privacy. The ever increasing privacy concerns in big data include knowing new and secret facts about people, combining their personal information with other data sets, adding value to their organizations with collected data from unaware people, threating illiterate people by predictive analysis of social media, tagging discriminated people by law enforcement, conflicting laws in different countries, lastly exchanging datasets between organizations [2]. To cope with such complex issues, laws and regulations must be enforced with clear-cut boundaries in terms of unauthorized access, data sharing, misuse, and reproduction of personal information.

V. Big Data Analytics for Security

Big data analytics aims to obtain beneficial information from large scale and complicated data [3]. The increase of stored or streamed data and development of analysis systems has led to using these activities in information security. The anomaly detection, intrusion detection, fraud detection, advanced persistent threats (APT) detection, and forensics from big data has been accomplished by examining the logs, system events, network traffic, website traffic, security information and event management (SIEM) alerts, cyber attack patterns, business processes and other information sources [24]. To detect these attacks, large volume and variety of data is accumulating and associate with network history. The advantageous uses of big data, such as performing without deletion of logs after a certain period, running complex queries on large and unstructured datasets, and facilitating human-computer interactions via visual interfaces, for security is becoming quicker and cheaper than traditional methods [24]. There is no need to delete the cancelled accounts or old logs as they can be used for the purpose of forensics later. In addition, real time and agile decision support applications, automatic defence and risk reduction systems, prediction of attack, determining of zero-day attack duration and tracking of attackers can be developed by analysing suspicious and malicious patterns from information security data [13,25,26].

A method to detect malware using big data has been proposed [27]. For this purpose, Large Iterative Multitier Ensemble (LIME) classifiers have been used to handle big data. The performance of LIME classifier with other base classifier was evaluated. The results showed that the presented method performed better than base classifiers. In another study, a framework with big data environment has been suggested such as big data analytic tool and NoSQL database for android application security assessment [28]. The white-box, black-box and mobile environment forensic approaches have been used to determine security assessment level. Then authors inserted assessment results into CouchDB, which is one of the NoSQL database. Using this database, they tried to discover security issues or visualization with big data analytic tools like Sckit, Matplotlib. Finally, they used SOA controller to publish their results via web service.

VI. CONCLUSION

Big data needs extra requirements for security and privacy in data gathering, storing, analysing, and transferring. In this paper, we examined studies on big data security and privacy, comparatively. According to the literature, network traffic should be encrypted with suitable standards; access to devices should be checked; employees should be authorized to access systems; analysis should be done on anonymised data; communication should be made for the secure channel to prevent leakage, and network should be monitored for threats.

Big data privacy, safety and security are the biggest issues to be discussed more in the future, so new techniques, technologies and solutions need to be developed in terms of human-computer interactions or existing technologies should be improved for accurate results. It is hoped that this study would help understand the big data and its ecosystem better and develop better systems, tools, structures and solutions not only for today but also for the future.

REFERENCES

[1] T. Vijey, A. Aiiad, “Big Data Security Issues Based on Quantum Cryptography and Privacy with Authentication for Mobile Data Center”, Procedia Computer Science, vol. 50, pp. 149–156, 2015.

[2] B. Matturdi, X. Zhou, S. Li, F. Lin, “Big Data security and privacy: A review”, Big Data, Cloud & Mobile Computing, China Communications vol.11, issue: 14, pp. 135 – 145, 2014.

[3] C.L.P. Chen, C.Y. Zhang, “Data Intensive applications, challenges, techniques and technologies: A survey on Big Data”, Information Sciences, vol. 275, pp.314-347, 2014.

[4] N. Miloslavskaya, M. Senatorov, A. Tolstoy, S. Zapechnikov, “Information Security Maintenance Issues for Big Security-Related Data”, Future Internet of Things and Cloud (FiCloud), pp. 361 – 366, Barcelona, 2014.

[5] Cloud Security Alliance Big Data Working Group, “Expanded Top Ten Big Data Security and Privacy Challenges”, April 2013.

[6] A.T.H. Ibrahim, Y. Ibrar, B.A. Nor, M. Salimah, G. Abdullah, U.K. Samee, “The rise of “big data” on cloud computing: Review and open research issues”, Information Systems, vol. 47, pp. 98–115, 2015.

[7] P. Adluru, S.S. Datla, Z. Xiaowen, “Hadoop eco system for big data security and privacy”, Systems, Applications and Technology Conference (LISAT), Long Island, Farmingdale, NY, pp. 1 – 6, 2015.

[8] M. Divakar, K. Shrikant, J. Shweta, IBM, “Big data architecture and patterns, Part 1: Introduction to big data classification and architecture”, http://www.ibm.com/developerworks/library/bd-archpatterns1/ (Accessed Date: 1 August, 2015).

[9] B. Saraladevi, N. Pazhaniraja, P. Victer Paul, M.S. Saleem Basha, P. Dhavachelvan, “Big Data and Hadoop-A Study in Security Perspective”, Procedia Computer Science, vol. 50, pp. 596 – 601, 2015.

[10] A. Kumar, L. HoonJae, R.P. Singh, “Efficient and secure Cloud storage for handling big data”, Information Science and Service Science and Data Mining (ISSDM), pp. 162 – 166, Taipei, 2012.

[11] H. Cheng, C. Rong, K. Hwang, W. Wang, Y. Li, “Secure big data storage and sharing scheme for cloud tenants” Communications, China, vol. 12, issue: 6, pp. 106 - 115, 2015.

[12] S. Marchal, J. Xiuyan, R. State, T. Engel, “A Big Data Architecture for Large Scale Security Monitoring”, Big Data (BigData Congress), pp. 56 – 63, Anchorage, AK, 2014.

[13] L. Liu, J. Lin, “Some Special Issues of Network Security Monitoring on Big Data Environments”, Dependable, Autonomic and Secure Computing (DASC), pp. 10 – 15, Chengdu, 2013.

[14] A. Gupta, A. Verma, P. Kalra, L. Kumar, “Big Data: A security compliance model”, IT in Business, Industry and Government (CSIBIG), pp. 1 - 5, Indore, 2014.

[15] L. Chang Liu, R. Ranjan, Y. Chi, Z. Xuyun, W. Lizhe, C. Jinjun, “MuR-DPA: Top-Down Levelled Multi-Replica Merkle Hash Tree Based Secure Public Auditing for Dynamic Big Data Storage on Cloud”, Computers, vol. 64, issue 9, pp. 2609 – 2622, 2015.

[16] H. Chingfang, Z. Bing, Z. Maoyuan, “A novel group key transfer for big data security”, Applied Mathematics and Computation, vol. 249, pp. 436–443, 2014.

[17] S. Junggab, K. DongHyun, R. Hussain, O. Heekuck, “Conditional proxy re-encryption for secure big data group sharing in cloud environment”, Computer Communications Workshops (INFOCOM WKSHPS), pp. 541 - 546, Toronto, ON, 2014.

[18] M.R. Islam, M.E. Islam, “An approach to provide security to unstructured Big Data”, Software, Knowledge, Information Management and Applications (SKIMA), Dhaka, pp. 1-5, 2014.

[19] T. Omer, P. Jules, “Big Data for All: Privacy and User Control in the Age of Analytics”, Northwestern Journal of Technology and Intellectual Property, article 1, vol. 11, issue 5, 2013.

[20] J. Sedayao, R. Bhardwaj, N. Gorade, “Making Big Data, Privacy, and Anonymization Work Together in the Enterprise: Experiences and Issues”, Big Data (BigData Congress), pp. 601 – 607, Anchorage, AK, 2014.

[21] J.P. Jisha, S.P. Anitha, “Adaptive Utility-based Anonymization Model: Performance Evaluation on Big Data Sets”, Procedia Computer Science, vol. 50, pp. 347 – 352, 2015.

[22] Z. Xuyun, L. Chang, S. Nepal, Y. Chi, “Wanchun Dou; Jinjun Chen, Combining Top-Down and Bottom-Up: Scalable Sub-tree Anonymization over Big Data Using MapReduce on Cloud”, Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 501 – 508, Melbourne, VIC, 2013.

[23] Z. Xuyun; D. Wanchun, P. Jian, S. Nepal, Y. Chi, L. Chang, C. Jinjun, “Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud”, Computers, vol. 64, issue 8, pp. 2293 - 2307,2015.

[24] A.A. Cardenas, P.K. Manadhata, S.P. Rajan, “Big Data Analytics for Security”, IEEE Security & Privacy, vol. 11 , issue 6, pp. 74 – 76, 2013.

[25] Cloud Security Alliance Big Data Working Group, “Big Data Analytics for Security Intelligence”, September 2013.

[26] T. Mahmood, U. Afzal, “Security Analytics: Big Data Analytics for cybersecurity: A review of trends, techniques and tools”, Information Assurance (NCIA), pp.129-134, Rawalpindi, 2013.

[27] J.H. Abawajy, A. Kelarev, M. Chowdhury, “Large Iterative Multitier Ensemble Classifiers for Security of Big Data”, Emerging Topics in Computing, vol. 2, issue 3, pp. 352 – 363, 2014.

[28] Z. Hongye, X. Jitian, Z. Xiangxin, “Enhance enterprise Android application security with cloud computing and big data analytics”, Information Technology and Artificial Intelligence Conference (ITAIC), pp. 238 – 243, Chongqing, 2014.