Secured System For Big Data Protection

1. Ekta Mandlik, Computer dept,MESCOE, Pune
2. Minal Pawar, Computer dept, MESCOE, Pune
3. Pooja Dhakane, Computer dept, MESCOE, Pune
4. Kajal Jambhale, Computer dept, MESCOE, Pune
5. Prof. Y.S. Ingle,Computer dept, MESCOE, Pune

***Abstract*—The maintenance of the big data in aData warehouserequires a plethora of attributes to be monitored such as, licenses, servers, etc. this is a highly time consuming and expensive approach. Therefore, most of the developers and organizations opt to outsource their Bigdata storage responsibilities. As the Third-party storage facilities have infrastructure and means for the maintenance of the Big Data. The integrity of the data can be affected by the employees of the organization or the third-party employees too who handle and maintain the integrity manually. Therehas been research done in this regard to automate most of this process through the handling of the tampered data tuples and its restoration. But these have been not up to the mark which is noticed by the increased time complexity that is observed in the restoration. Therefore, the presented technique has been outlined in this regard that handles the Big Data stored in the Data warehouse automatically. The system performs this task through the Avalanche effect identification through recursive bilinear pairing. The notarization and validation are implemented to identify the attributes of the tamper such as whichand When.**

Keywords:*Database Integrity, Bilinear Pairing, Avalanche Effect, Validation, Notarization.*

I. INTRODUCTION

Data can be defined as the quantities, characters, or symbols or its operations on a unit performed by a laptop computer, which might be kept and transmitted through the variability of electrical signals and recorded on magnetic, optical, or mechanical recording media.

Big knowledge is additional knowledge that is generated through a large amount of data or massive knowledge that is the term used to describe a set of knowledge that's immense in size and nevertheless growing exponentially with time. briefly, such knowledge is thus massive and complicated that none of the standard knowledge management tools square measure able to store it or method it with efficiency.

There are three types as follows like Structured ,Unstructured and Semi-structured

Structured - Any information that may be held on, accessed and processed within the kind of varied format is

termed as 'structured' information. Over a large amount of time, talent in technology has achieved bigger success in developing techniques for operating with such reasonable information (where the format is accepted in advance) and conjointly explaining worth out of it. However, nowadays, we tend to foresee problems once the size of such information grows to an enormous extent, typical sizes being within the rage of multiple zettabytes.

Unstructured -Any information on the unknown kind of unknown structure is classed as unstructured information. to convert to the size being Brobdingnagian, un-structured information poses multiple challenges in terms of its method for the derivation of useful information out of it. A typical example of unstructured information could also be a heterogeneous information offering containing a mix of simple text files, images, videos, etc. present-day organizations have wealth of knowledge on the market with them but sadly, they do not have algorithms powerful enough to derive the worth out of it since this information is in its raw kind or unstructured format.

Semi-structured -Semi-structured knowledge will contain each variety of knowledge. we are able to see semi-structured knowledge as a structured data but however, it's really not outlined with e.g. a table definition in relative DBMS. An example of semi-structured knowledge may be a knowledge described by an Associate Nursing XML file.

The GDPR applies to the "processing" of "personal data". As these definitions and the interpretation are interpreted as terribly broad, varied obligations beneath the GDPR can apply in several circumstances once handling huge information and big data analytics are used.

Moreover, within the context of huge information, it cannot be excluded that analysis should have considerations towards "sensitive data" – the process of that is restricted and prohibited in most cases – or that it'll have a “transformational impact” on data. as an example, the process of non-sensitive personal information may lead to, leakages of personal data – through data processing, as an example that reveals sensitive information concerning a person.

The broad scope of application of the GDPR is usually implemented and therefore the attainable process of sensitive information could need limiting. befit process activities or

technical developments to tackle the demanding rules enclosed within the GDPR.

In case personal knowledge is being processed (as it's the case in knowledge analytics), it's necessary to look at the concrete scenario thus to confirm exactly the precise role performed by the various actors concerned in such processes. The assorted ideas enriched beneath EU knowledge protection law and especially the distinction between "data controller" and "data processor", in addition to their interaction, is of preponderating importance so as to see the responsibilities. within the same vein, such ideas are essential so as to see the territorial application of information protection law and therefore the competency of the superordinate authorities.

The qualification of actors and therefore the distinction between "controller" and "processor" will quickly become complicated during a massive knowledge context. this can be very true taking under consideration extra knowledge protection roles like joint-controllership, controllers in common, and sub-processors. this can be done, chiefly thanks to the very fact that several actors could also be concerned within the knowledge within the chain, the mapping of which may be rather heavy.

Hence, extra steerage and model agreements, compliant with the strict necessities of the GDPR, which are welcome to clarify the relationships within the massive knowledge cycle.

- The GDPR outlines six information protection principles one should consider in detail once before processing personal information, most of that is being challenged by some key options of huge information.

• The principle of "lawfulness" implies every process of private information ought to be supported by a legal ground (see next section).

• The principle of "fairness and transparency" means the controller should offer information to people regarding its processing of their information unless the individual already has this info. The transparency principle during a huge information context – wherever the complexity of the analytics renders the process opaque – will become notably difficult and implies that "individuals should learn clear information on what information is processed, together with information determined or inferred regarding them; however and for what functions their info is employed must be studied, together with the logic utilized in algorithms to see assumptions and predictions regarding them."

Consumers' brands did face a tough leveling act as they require to deliver extremely customized services and solutions. massive information makes that attainable.

However, analytics technology is changing into a more proficient system that is good at distinguishing shopper identities. whereas most brands don't track the identity of every one of their users, they usually track enough data to guess someone's identity.

Sympathy for privacy rights is additionally wearing away, because of the proliferation of internet trolls and online abusers. High-profile incidents, like Gamergate doxing and harassment and also the cyberbullying of Amanda Todd, has crystallized opinions of several lawmakers and influencers to argue for additional restrictions on online privacy.

These trends indicate that anonymization might shortly become an issue of the past. this could concern online users that need to shield their online privacy. it's one among the explanations folks deliver for victimization VPNs.

Growing risk of security breaches

Customers are terribly involved in cases regarding security breaches currently. Breaches at Target, Yahoo and alternative major corporations indicate that even multinational brands have facilities that are usually too lax with their security protocols. The target according to that regarding one hundred ten million users that were plagued by their security breach in 2014. This is a grave downside, as a result of this the brand's store very sensitive information on their customers. Their client records contain Mastercard numbers, Social Security numbers, addresses and lots of alternative terribly personal items of knowledge.

As last year's White House Report indicates, massive information will produce the danger of unintentional discrimination. disposition actuaries, employers, faculty admissions officers, and alternative decision makers believe there will be a terribly heavy toll on analysis centers on massive information to form key choices. Since their information includes info on client demographics, brands could unknowingly develop algorithms that punish individuals supported quality, gender or age.

Brands have become a lot of and a lot of hooked into big data information currently. sadly, they typically have an excessive amount of useless information in their algorithms and therefore the accuracy of their information.

The approaches that brands use to gather information could also be blemished from the start, which implies they will collect inaccurate info. information can also be compromised by hackers, malware, disk harm, and alternative problems. If brands produce models around inaccurate information, it will cause serious complications for all users concerned. for instance, insurance actuaries could produce inaccurate risk profiles supported blemished information, which can provide a blow below the belt to penalize each user in their cluster.

This research paper dedicates section 2 for analysis of past work as literature survey and Section 3 works on proposed methodology details. Section 4works on the evaluation of the results and finally section 5 concludes this paper with the hints on future scopeenhancement.

II. LITERATURE SURVEY

This section of the literature survey eventually reveals some facts based on thoughtful analysis of many authors work as follows.

Y. Canbay [1] introduces Big data concept which introduced in 2012 and it was described as "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and

process automation" The data which cannot be managed by the long-established framework and system is not able to handle it in terms of volume, velocity, and variety it is termed as big data. Big data brought many things such as great opportunities, activities such as policymaking, strategy development, selection management, future planning, forecasting. As it is newborn technology it faces various problems such as terms of complexity.

K. Saravanan [2] defines a large collection of meaningful information which is also known as Big data. Conventional database systems cannot handle these types of data sets. The data can be of any type such as text, audio, video, etc., are generated by different streams. Every day massive data is generated. In big data there two things are very important, the security of data and privacy of data. Thus, security and privacy are two different things security is safeguarding the data from attackers and abuse of data whereas privacy is focused on the apparent person whose data is gathered, shared and used in the right perspective way. Thus, the proposed paper explains that big data privacy protection strategies of various stages have been analyzed.

H. Cheng [3] explains that Big data is the future of information technology architecture, along with cloud computing which has been an emerging topic to be widely researched. There are some challenges such as data privacy protection, cloud tenants concern, and conventional protection technology. Different Cloud service providers (CSPs) put huge computing processes and run on various computing resources. In the proposed technique data privacy in cloud computing is achieved through encryption. Hash operations are performed to protect the data privacy which will be stored in the cloud. The proposed scheme can protect the tenants' data privacy efficiently.

B. Jayasingh [4] elaborates that the issues of cybersecurity in terms of finding the attacker can be solved using a big data environment. The issues such as fraud detection, network forensics, data privacy issues, and data provenance problems are well suited for the proposed methodology. The authors concentrate on tools and techniques of data mining in terms of security and the use of these techniques for security to protect big data encryption capabilities is used. In the proposed paper they have used the Bayesian classification algorithm for classification and data mining is suitable for predicting the attack type.

F. Liang [5] discusses the latest technology such as mobile and social networking applications, and Internet of Things (IoT)-based smart-world systems, smart grid, smart transportation, smart city, and others, where massive amounts of data will be collected. The main reason behind huge data and huge datasets is different kinds of sensors and smart devices collectively generate a large amount of data. The main purpose of the paper is to provide a clear and deep understanding of big data trading. The estimated data generated every day is 2.5 quintillion bytes. The proposed paper evaluated digital copyright protection mechanisms, including Digital Copyright Identifier (DCI), Digital Rights Management (DRM), Digital Encryption, Watermarking, and others, and outline challenges in data protection in the data trading lifecycle.

L. Xu [6] narrates that the data mining technologies are one of the fastest-growing technologies in recent times emerging research topic in data mining, known as privacy-preserving data mining (PPDM). PPDM mainly concentrates on how to bring down the privacy prospect brought by data mining, which causes undesired declaration of sensitive information that may also appear in the process of data collecting, data publishing, and information delivering. The authors identify four different types of users in data mining applications data provider, data collector, data miner, and decision-maker. How to protect sensitive information from the security threats brought by data mining is proposed in this paper.

K. Gai [7] declares numerous applications in people's life recently is only because of big data and cloud computing techniques. As this is the emerging technique cloud computing has extended into numerous fields so that many new services can be introduced such as mobile parallel computing. The proposed system concentrates on the privacy issue of big data and cloud computing to solve the issue of such as maximize the efficiency of privacy protections D2ES approach is designed. The results show that the proposed approach had adaptive and superior performance in comparison to conventional techniques.

T. Basso [8] estimates that protecting individual privacy is one of the most challenging issues in the context of Big Data. The rise of the Internet of Things (IoT) increases this effect, with the increasing number of wireless sensor networks providing data about many areas of our lives. It provides the potential to analyze and control both natural resources and urban environments, Creates huge volumes of data coming from different origins, e.g., sensors, devices, networks, log files, transactional applications, social media, etc. Thus the development of tools for privacy breach detection is a much-needed addition in this direction.

Y. Huang [9] narrates that the data rinse is an extensive issue where real data is rarely error-free. Current approaches in networking and cloud infrastructure have prompted a new computing paradigm called Database as-a-service. In the proposed paper the author presents PACAS: a Privacy-Aware data Cleaning-As-a-Service technique that eases communication within the client and the service provider via a data pricing scheme where the client fires the queries, and the service provider returns clean answers and solution for a price while preserving private data. PACAS has demonstrated to successfully safeguard semantically related sensitive values,

and provide upgraded accuracy over existing privacy-aware cleaning techniques.

C. Yin [10] explains that devices in the Internet of Things (IoT) paradigm originate, process, and exchange huge amounts of security and safety-critical data as well as privacy-sensitive information, and hence are attracting different attackers and being targets of different types of attacks. Data gathered, aggregated and transmitted in sensor networks carry personal and sensitive information, which directly or indirectly discloses the state of a person. The author proposes location privacy protection which depends on privacy plans for the sensor network. The methodology is more effective and preserves the privacy of the data better than most conventional techniques.

Y. Qing [11] proposes that china is made fast growth in the field of data in comparison to the rest of the global in very few years. Big data is an emerging strategic resource, a new"petroleum" for the incoming years, and the main engine that will drive the socio-economic development. China is the world's second-wealthiest country is facing pressure from various provenances such as economic restructuring and improving, public services betterment and enhancement, and continued promotion in environmental preservation. The main aim is to develop efficient data storage security standards. There is a need to supervise and control equipment quality to make sure the data storage volume of the big data server is consistent, and to restrict data loss phenomenon introduced by data server compromise, at the same time, upgrade hard disk failure prediction technology, and host data transfer before a failure occurs.

J. Sedayao [12] estimates one of the largest concerns that any agency has with Cloud Computing, particularly in public Clouds, is preserving the secrecy, of data. Anonymization would appear to solve this issue by making it so that data could be handled without concern as to who looked at it. Human Factors Engineering group wanted to use web page access logs and Big Data tools to upgrade the usability of Intel's heavily used internal web portal. Big Data methods could produce benefits in the enterprise environment even when waged on anonymized data. The author concludes that Big Data, anonymization, and privacy can be hopefully combined but there is a need for observing data sets to make sure that anonymization is not at risk to correlation attacks.

K. Liu [13] discusses big data analytics that has widely been used in diverse areas where security is a crucial matter in big data surveys on personal data is the possibility to be a leak of personal privacy. It is mandatory to have an anonymization-based de-identification method to keep away from nasty privacy leaks. This kind of technique can prevent published data from being traced back to uncover personal private data. A prior empirical investigation has provided a method to lessen the privacy leak risk, e.g. Maximum Distance to Average Vector (MDAV), Condensation Approach and Differential Privacy. The proposed technique is more reliable in creating anonymous data and reduces the information loss rate.

Y. Xia [14] explains in the era of big data there are massive amounts of personal and company data that can be effortlessly collected by third parties like national statistical agencies, survey organizations, hospitals, credit card companies, and social media. Liberating the data and sharing the probability information of the data can bring much better information to policymakers, the national economy, and society. Two often-used data masking plans of action are the Additive Noise Data Masking Scheme and Multiplicative Noise Data Masking Scheme. This technique provides a new data mining technique for big data when data privacy is the main concern.

K. Abouelmehdi [15] introduces that Big data can unleash new opportunities for gathering knowledge for economic and social sciences and gives data a form of business value. Basically, the big data depends on the four v Volume, Velocity, Variety, Veracity. And the supplementary dimension that is examined is Complexity. The leading HDFS components include NameNode, DataNode, and BackUp Node. The technical probability is of big data is a cause of concern as the visibility requirement provided by the legislation on data protection is mandatory and can effectively preserve the privacy of the users. Thus, the Author proposes the various risks appearing in Big Data that can be made use by attackers to gain an unfair advantage.

III. PROPOSED METHODOLOGY

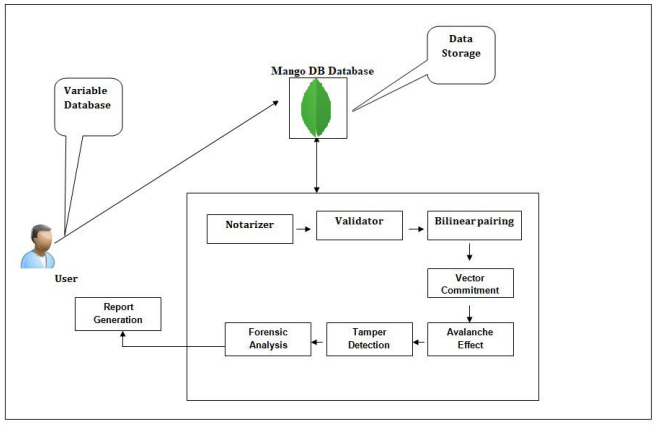

Figure 1: System Overview

The presented systemfor a variable database integrity maintenance can be illustratedthrough the procedure in figure 1. And these steps are outlined in detail in the steps mentioned below.

Step1: Notarizer – The Notarizer is the initial step, in this step the third-party organization which is tasked with the storage of the database gets a unique notarization key based on their attributes. The MD5 hash key generation is utilized to create through the random selection of seven characters, based onthe attributes of the client. When the client is uploading the data to the cloud this key is utilized to authenticate the client effectively and securely.

Step 2: Validator –The data once stored in the third-party servers is subject to a validation process for the designated time period such as 1,2,3, or 60-minute time period. The data vectors for the previous and the current data is preserved for the purpose of the determining any Tampering on the data.

Step 3: Bilinear Pairing and Avalanche Effect –This is the step where the actual verification of the data integrity of all the database tuples is performed. This is done through the generation of the hash key pairs on the designated validation time period defined in the previous step. These pairs of hash keys are referred to as the bilinear pairs. These pairs are utilized for performing the integrity evaluation. This is done by comparing the pairs, this is due to the fact that a minute change on the data would drastically alter the resultant hash key through the Avalanche effect. The identification of the tampered data tuples is detection of any tampering done on the tuples; the tampered ID is generated using the primary key of the tuples. This tampered ID illustrates the targeted tuples in the database that are subjected to an attack by the database attacker. This procedure is detailed in the Algorithm 1 given below.

After the generation of the tampering ID it is established through the avalanche effect that data has been compromised. Therefore, an extensive evaluation of the rest of the database is performed, this is done by comparing the hash keys of the current as well as the previous thread of the bilinear pairs that are generated. This process is repeated for all of the data in the database to reveal the extent of the tampering process.

The presented technique is detailedin the below Algorithm 1.

Algorithm 1: Tamper Detection using Bilinear Pairing

// Input : Database D_B, Attribute List AT_L, Time T
// Output : Tamper Report List TP_R
Function :tamper_detection(D_B , AT_L)
0: Start
1: PV_{LIST}=Ø, CV_{LIST} =Ø
[$PV_{LIST:}$ Previous DB List , $CV_{LIST:}$ Current DB List]
2: count=0
3: ***while TRUE***
4: ***wait for*T**
5: ***if*** (count=0), ***then***
6: CV_{LIST} = getDB (D_B)
7: PV_{LIST}= CV_{LIST}
8: ***end if***
9: ***else***
10: CV_{LIST} = getDB (D_B)
11: ***for*** i=0 **to** size of CV_{LIST}
12: R_{L1}= $CV_{LIST[i]}$ [ROW LIST]
13: H_{K1}= hashkey(R_{L1}) [Hash Key]
14: ***for*** j=0 **to** size of PV_{LIST}
15: R_{L2}= $CV_{LIST\ [j]}$ [ROW LIST]
16: H_{K2}= hashkey(R_{L2}) [Hash Key]
17: ***if*** ($H_{K1} \neq H_{K2}$), ***then***
18: check AT_L for Details
19: Generate Report $\mathbf{G_R}$
20 : $TP_R = TP_R + \mathbf{G_R}$
21: ***end if***
22: ***end for***
23: ***end for***
24: PV_{LIST}= CV_{LIST}
25: ***end else***
26: count++
27: ***end While***
28: ***return*** TP_R
29: Stop

Step 4: Tamper Detection and Forensic Analysis –This is the final step of the system where the action is performed extensively on the database to extract the relevant information on the tampering of the database. The extracted information answers the questions such as the time,date andattributes names of tampering. These attributes once determined are used to generate a detailed report which is shared with the admin.

After the generation of the report the previous string of the bilinear pair is restored in the database for the particular tampered ID all the other attributes are updated to get the original database tuples.

IV. RESULT AND DISCUSSIONS

The presented technique of Variable database integrity maintenance system is developed on the NetBeans IDE and Mongo DB as the database utilizing the Java programming language. Presented model utilizesawindows-based laptopequipped with a Core i5 processor and 6GB of primary memory. Intensive experiments are conducted to calculate the impact of the proposed methodology with thebelow mentioned approaches.

RMSE is calculated as the error difference between the obtained and the expected values. Which specificallyenables the most insightful results for the database tamper detection

system. And RMSE can be calculatedthrough the following equation 1.

$$RMSE= \sqrt{(xp - xo)}^2$$

Where
Xp –Expected number of tampered tuples
Xo –Obtained number of tampered tuples

The Result of the RMSE is completely unexpected for the presented model as it identifies most of the tampered tuples and gives a comprehensive report as it is evident through experimental data is tabulated in the below given table 1. Table 1 illustrates the RMSE of the proposed system is zero as it identifies all the possible details related to the tampered data providescomprehensive results. This depicts the efficiency of the proposed methodology where it attainsextraordinaryoutcomes for the variable evaluation of the tampered data in various different scenarios of tampering.

Experiment No	Total no of Data in Database	Total No of Tampered Data	Total no of Correctly Detected Tampering	RMSE
1	100	9	9	0
2	200	11	10	1
3	300	13	12	1
4	400	16	15	1
5	500	19	18	1

Table 1: RMSE Measurement Table

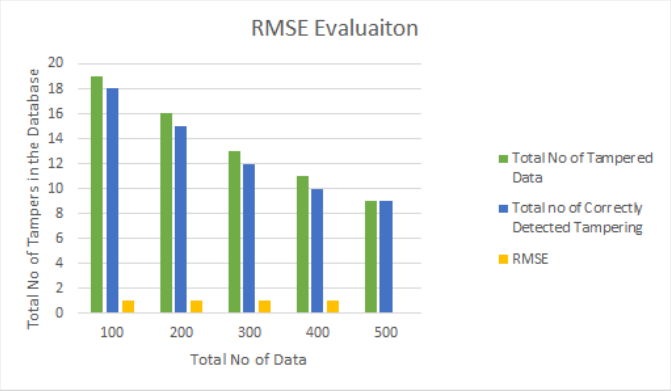

Figure 2: RMSE evaluation Results

The analysis of the above graph and table spill some light on the proposed model's outcomes. According to the experiments the proposed model achieves an excellent RMSE of 0.89 towards the error in RMSE in the tamper detection process. The yielded RMSE value indicates that the proposed model is working right in the entiremanner in the first attempt of deployment for securing the big data.

V.CONCLUSION AND FUTURESCOPE

This publication article deals with the implementation of a database integrity maintenance and protection of the variable database against tampering in bigdata. The presented technique implements the notarization approach that provides a unique key to the third-party vendors. The validator then validates the user to attempt any procedures to be performed on the stored data. The stored data is verified for its integrity using the Bilinear paring approach. The hash keys that are generated through the use of the MD5 algorithm are extracted from the tuples according to a time period. The keys are in pairs, such as the current and the previous keys which are called as the bilinear pairs. These keys are compared to notice any indications of an avalanche effect to identify any tampering that has been done on the tuples. When the tampering has been identified, a tampering ID is extracted from the primary key of the database tuple. The respective database is then effectively analyzed for the tampering done on the entire database. The previous string of the bilinear pair is then utilized to restore the original database tuples on the tampering ID generated. The RMSE approach has been utilized to provide an experimental analysis of the methodology which results in an impressive score of 0.89.

This presented technique can be improved in the future for the big data platform to execute in real time utilizing distributed computing on remote database servers in the cloud storage.

REFERENCES

[1] Yavuz Canbay, Yılmaz Vural, Seref Sagiroglu, "Privacy-Preserving Big Data Publishing", International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, 2018.

[2] Ms. D. Viji," A Journey on Privacy protection strategies in big data " International Conference on Intelligent Computing and Control Systems ICICCS, 2017.

[3] Hongbing Cheng, Weihong Wang, Chunming Rong," Privacy Protection Beyond Encryption for Cloud Big Data" 2nd International Conference on Information Technology and Electronic Commerce, ICITEC 2014.

[4] Bipin Bihari Jayasingh, 2M. R. Patra, 3D Bhanu Mahesh, "Security Issues and Challenges of Big Data Analytics and Visualization" 2nd International Conference on Contemporary Computing and Informatics (IC3I), 2016.

[5] Fan Liang, Wei Yu, Dou Any, Qingyu Yang, Xinwen Fux, and Wei Zhao, "A Survey on Big Data Market: Pricing, Trading, and Protection", IEEE Access, 2018.

[6] Lei Xu, Chunxiao Jiang Jian Wang, Jian Yuan, and Yong Ren, "Information Security in Big Data: Privacy and Data Mining", IEEE Access, 2014.

[7] Keke Gai, Meikang Qiu, Hui Zhao, and Jian Xiong, "Privacy-Aware Adaptive Data Encryption Strategy of Big Data in Cloud Computing", IEEE 3rd International Conference on Cyber Security and Cloud Computing, 2016.

[8] Tania Basso, Roberta Matsunaga, Regina Moraes, and Nuno Antunes, "Challenges on Anonymity, Privacy, and Big Data", Seventh Latin-American Symposium on Dependable Computing, 2016.

[9] Yu Huang, Mostafa Milani, Fei Chiang, "PACAS: Privacy-Aware, Data Cleaning-as-a-Service", IEEE International Conference on Big Data (Big Data), 2018.

[10] Chunyong Yin, Jinwen Xi, Ruxia Sun, and Jin Wang, "Location Privacy Protection based on Differential Privacy Strategy for Big Data in Industrial Internet-of-Things", IEEE Transactions on Industrial Informatics, 2017.

[11] L. Yue, H. Jia, G. Minji, Y. Qing, and Z. Xinsheng, "An Overview of Big Data Industry in China", China Communications • December 2014.

[12] Jeff Sedayao, and Rahul Bhardwaj, "Making Big Data, Privacy, and Anonymization work together in the Enterprise: Experiences and Issues", IEEE International Congress on Big Data, 2014.

[13] Kai-Cheng Liu, Chuan-Wei Kuo, Wen-Chiuan Liao, and Pang-Chieh Wang, "Optimized data de-identification using multidimensional k-anonymity", 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering, 2018.

[14] Yan-Xia Lin, "Mining the Statistical Information of Confidential Data from Noise-Multiplied Data", IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress, 2017

[15] Karim Abouelmehdi, Abderrahim Beni-Hssane, Hayat Khaloufi, and Mostafa Saadi, "Big Data Emerging Issues: Hadoop Security and Privacy", 5th International Conference on Multimedia Computing and Systems (ICMCS), 2016.

*****