International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Security Methods for Privacy Preserving and Data Sharing Over Cloud Computing and Big Data Frameworks

Kapilesh S. Swami[1], Dr. P Sai Kiran[2]

[1]Research Scholar, Department of Computer Science and Engineering, K L University, Vijayawada, India

[2]Director - Coding Club, Department of Computer Science and Engineering, K L University, Vijayawada, India

Abstract: ***The cloud computing is one of the widely used services for resource management by many IT (information technology) and non-IT organizations due to its different benefits in terms of time saving and cost savings to the companies. Such cloud computing frameworks are used to store the small to big data efficiently. Most of companies want to store huge amount of data and hence along with cloud computing it is important to have big data platforms to handle big data operations. The term big data is nothing but the huge amount of semi-structured, unstructured and structured which is having the ability of such huge information processing. Whereas cloud computing is technology which is used to provide the valuable IT services in term of resources like software, infrastructure, platform, storage etc. For both big data and cloud computing, data storage is main goal of any organization. As cloud computing is open environment technology, security is major challenge for each organization while storing their important data over the clouds. Therefore, security of data and privacy preserving is gaining the significant attention of many individual or group of researchers for both cloud computing and big data frameworks. In literature, number of security techniques introduced for data security and privacy preservation in cloud computing. The goal of this paper is to present the study over all recent data security and privacy preserving methods for cloud computing as well as big data frameworks with its comparative analysis.***

Keywords: Cloud Computing, Encryption, Decryption, Privacy Preserving, Big Data, Cryptography, Data Sharing, Data Storage.

1. Introduction

Concept of cloud computing is nothing but the set of available resources those are assigned based on end users demands. This defines the novel approach of providing the services. Such innovating approaches of services areoffering the number of advantages to the small to large organizations for their important data management. Cloud Computing is one of the matchless and having the new label to the traditional idea. The collection of resources in cloud computing framework is provided by service provider of cloud to the end users based on their demands through the internet. The services of cloud are distributed all over the world. Cloud computing creates the virtual environment for its users in which it allowing the end users to use services or resources of cloud virtually. Due to number of advantages of using cloud computing framework, it becomes spotlight in just few years of span. The common example of cloud services is Microsoft office 365, Google Search Engine, Oracle Cloud etc. Along with the number of benefits of using cloud computing services, there are number of challenges also such as availability, reliability, and most important and widely studied research problem is security. In this paper we are focusing on study over various security methods. The growing use of cloud services resulted into number of security challenges. Absence of data security method is main concern in traditional cloud computing services and platforms.

On other hand, many organizations used the big data platforms processing, organizing and analysing the huge amount of data before storing to cloud computing. Hence with the use of both big data and cloud computing, security methods play very significant role in order to make sure that entire data should be secure. The challenge of data security as well as privacy preserving is lifted by variety, volume, and velocity of big data frameworks. The diversity in data sources, multiple domains generated data are combined together in order to form huge amount of data which leads to the many security concerns. As the benefits of using big data frameworks, since from few years most of companies started using the big data platforms in order to process large volume data. Hence the requirement of cloud computing framework and big data frameworks is increasing now days due to the significant growth digitization of data in many organizations. This leads to the generation of structured, semi-structured as well as unstructured data. The methodology of collection of data, its analysis, sorting and then mining is leading to collection of huge amount of individual user's sensitive information. Such type of information is satisfying the enterprises demands and delivering the services to many organizations in case this data is stored into the big data platform [1].

The earlier servers of cloud services were storing only data which is either plain text or encrypted formation. Such stored data over the cloud is treated as dead as it was not provided to perform any processing or analysis of such stored data on traditional cloud servers. However, in recent times, it becomes possible to perform processing on stored data over cloud servers with the help of using big data systems. Big data systems are providing the huge amount of data storage at cloud as well as supporting the computational operations. Here the computation operations are data processing, conversion, data encryption, analysis etc. This operation is performing to refresh dead data on cloud servers [2] [3].

Data security and confidentiality are main requirements such cloud computing systems with big data frameworks. For data security and privacy preserving, there are number of different security methods introduced by various research groups. Some these methods are based on cryptography approach with goal of achieving the data security under cloud computing systems. The method Public Key Encryption (PKE) for example allows data encryption at sender side by using the public key of intended receiver of data with goal of providing the data security and it should be accessed by legitimate users only. But the limitation of this approach is that it failed to address all the requirements of end users for cloud data security.

In hospital day to day operations in which the medical information of individual patients is stored into cloud storage systems. The data stored over the cloud system is in form of encrypted using PKE approach with goal of preventing the invalid access to patient's sensitive data by invalid user. Therefore, after storing patient's data in encrypted format over the cloud server, only legitimate users can able to access or having privileges for accessing such data from cloud server. The recently introduced methods along with PKE such as identity based encryption (IBE) as well as attribute based encryption (ABE) achieving the confidentiality of such medical data efficiently [4] [5].

The earlier methods of cryptography those are introduced for guaranteed confidentiality of personal medical information's are failed from preventing the information leakage. This is because as existing cryptography techniques not considering the anonymity of sender or receiver ciphertext. Hence it is possible that novice or invalid user can able to obtain the information of ciphertext. For example, cloud server which may know the information of public key used for encryption, and hence it may possible for cloud server to obtain the ciphertext information with name of patient related to that ciphertext. Similarly, recipient of ciphertext may be identified from the available ciphertext without any trouble. And hence this is the serious cause to the privacy of patient [6].

It is observed that current techniques of security systems for data sharing are able to solve the problem of security and privacy preserving research problems. These methods are solving such problems partially only by considering the various views of data storage. Therefore, these methods do not achieve the complete life cycle of data security under the cloud computing environment [7]. Suh existing techniques are accepted partially for data storage in cloud computing, but if there are big data systems presents for data processing in which data from multiple domains and stakeholders are combined together, it will be the serious cause to avoid any security risk which may leads to the loss of user's important information. Hence it is required to have complete security framework by considering the involvement of big data systems for data sharing as well as privacy preserving. This framework should consider both efficiency as well as all the security concerns of data storage and sharing. We studied recently the method in which author proposed the method for sensitive data sharing by addressing the high level of security on big data systems [9]. This method providing the secured delivery of data, secure data usages, secure storage and destruction of data securely. The base of such method proxy re-encryption technique which is used to address the research challenges related to data security and privacy preserving based on the VMM (virtual machine monitor). This method achieving the efficient security solutions by considering all factors [9]. But the limitation of this technique is that it is suffering from the issues of efficiency in terms of processing overhead, encryption timing as most of operations are executed on sender side as well as server side. Another problem of this method is that it cannot solve the problem of user anonymity [9].

The goal of this research article is to present the complete study on cloud computing and big data systems security methods along with its challenges and concerns. This paper presenting comparative study among recently presented existing methods for data security and privacy preserving with aim of listing the current research challenges to address in domain of security. The future roadmap is presented by this paper for our upcoming research studies. Section II discussing the review over the cloud computing growth and security concerns based on survey of IDC. Section III presenting the study over current challenges and issues cloud computing security. Section IV, presenting the recent methods of cloud computing and big data security methods. Section V, presenting the comparative study and listing the limitations of current systems of security. Section VI presenting the conclusion and future work details for this research work.

2. Cloud Computing By IDC

In this section, the survey over the cloud computing which is conducted by IDC (international data corporation) is presented. This study presents the benefits and growth of cloud computing framework which used by different kinds of industries along with big data platforms. In below paragraphs, we are presenting study over the cloud computing growth in different ways, security survey, current and future usages detail of cloud computing etc. [10 [11].

- Growth of Cloud: Below table 1 is showing the growth ratio performance of cloud since from 2008 to 2016. From this table it showing that performance of usage of cloud services by different industries is growing from 2008 still to present.
- Cloud Security Analysis: Figure 1 is showing the analysis of security over cloud systems in different aspects. This study shows that security as the first rank as per the IT executives. The information showing in figure 1 over the security which is based on data collected from 300 professionals by asking questions of cloud security and systems.
- Top Technology Priorities: The figure 2 is showing the survey on first 10 technologies which is widely used by companies since from last 10 years. From this figure it is clear that cloud computing is at first rank in usage.
- Cloud Services Revenue: From the report generated by IDC, we observed that in 2009, revenue of cloud is around 17.4 billion dollars, whereas same it was 44.2 billion dollars in 2013.
- Cloud Usage: Figure 3 is showing the analysis of present usage of cloud services in specific industry and next three years' usages for the same industry. It is showing that

current usage and future usage having more difference as it is predicted that future usage will be more.

Table 1: Cloud usage growth performance by ITC

	2008	2012	2016	Growth (%)
Cloud IT Spending	$16 B	$42 B	$80 B	45%
Total IT spending	$383 B	$494 B	$612 B	28%
Total-cloud spend	$367 B	$452 B	$590 B	27%
Cloud Total spend	4%	9%	17%	

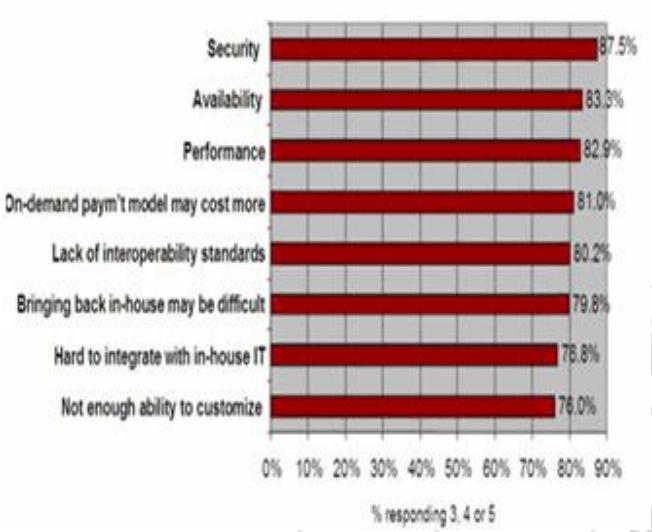

Figure 1: Review of cloud security system

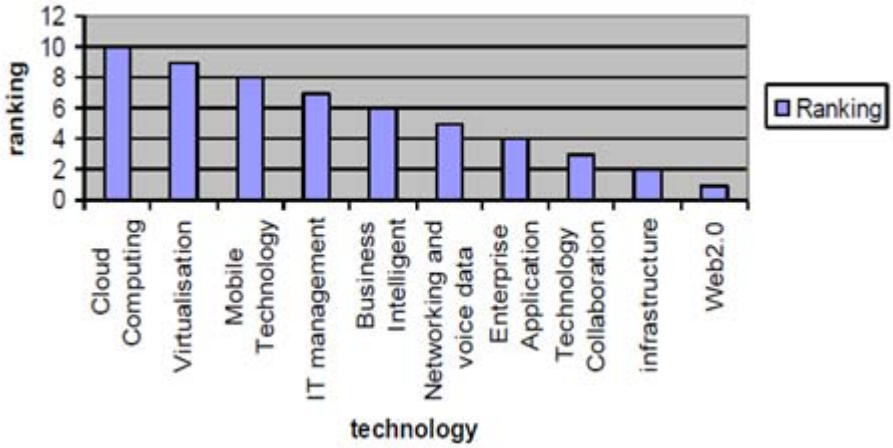

Figure 2: Priority chart for top 10 technologies

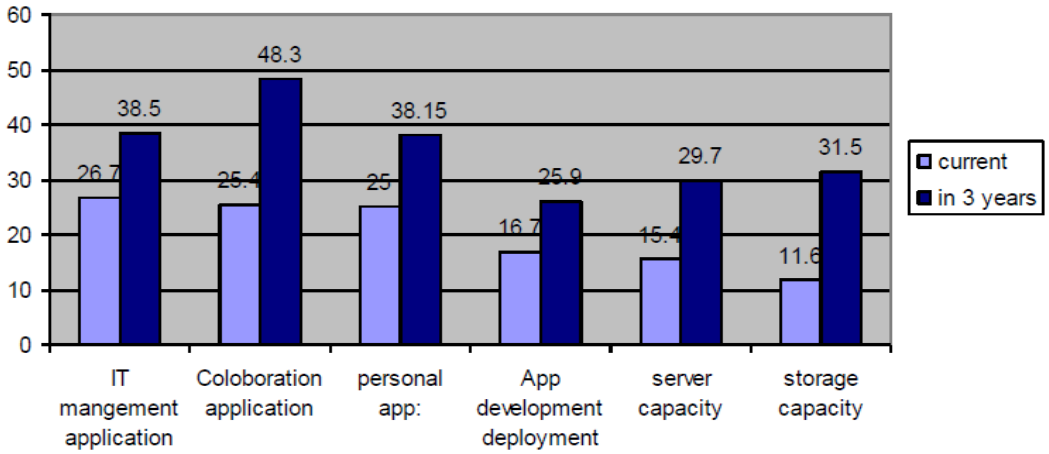

Figure 3: Present and future cloud storage report by IDC

3. Security Challenges and Issues

As we studied that cloud computing is one of the emerging approach along with the less cost, shared resources, pay according to the use based on end user demand. Because of number of characteristics it is impacted on budget in IT as well as issues related privacy, security and security challenges. The goal of this section is present security challenges and issues in cloud computing.

Under cloud computing environment, end user not aware about where is data stored, who is manging that data as well as other vulnerabilities in cloud computing environment. Below listed are common issues those can be faced by cloud service provider during the implementation of cloud services.

3.1. Privacy Related Issue

Securing the individual private information is particular individual's right. Privacy in cloud computing framework is based on model of cloud deployment. The public cloud is one of the promising architecture by considering the concern related to the reduction of cost. However, this public cloud is relying over the cloud service provider in order to manage as well as keep end user's information which leading to number of privacy related issues such as:

3.1.1. Lack of User Control: In cloud computing framework, processing and sharing of user's important information is not having enough user control while leads to the threats like theft, misuse, or illegal access [11].

3.1.2. Unauthorized Secondary Usage: The data security and profit in cloud computing frameworks is achieved by placing data to legitimate and authorized secondary uses. At present, there are not technical barriers for the secondary uses.

3.1.3. Transborder Dataflow and Data Proliferation: Data proliferation is attributing of cloud computing systems which is composed of number of companies and hence not managed as well as controlled by individual data owners. Copying the data over the many datacentres guarantees the ease of use. It is very challenging to make sure that backup of data or its duplicate copy is not saved or processed with specific authority.

3.1.4. Dynamic Provision: Nature of cloud is vibrant therefore it is not clear that who is responsible legally in order ensure the sensitive data privacy which is stored by end users on cloud server.

3.2. Security

In cloud computing paradigm, public cloud is not only increasing the issues of privacy but also increases the security concerns. Below are listed are common security concerns for public cloud:

3.2.1. Access: Accessing the personal and private information is one the threat to cloud security. It may possible that any attacker can try to access the personal information.

3.2.2. Data Lifecycle Control: To ensure the customer that it has control over data, if it remove or delete data vendor cannot regain this data.

3.2.3. Availability and Backup: There is no any surety of availability and back up of data in this environment. In business backup is one of the important consideration.

3.2.4. Multi-Tenancy: It is feature of SAAS that one program can run to multiple machines. CSP use multi-tenant application of cloud to reduce cost by using virtual machine but it increases more vulnerability.

3.2.5. Audit: To implement internal monitoring control CSP need external audit mechanism. But still cloud fails to provide auditing of the transaction without effecting integrity.

3.3. Trust

For every organization, trust is required to gain the profit by using the cloud services. However, cloud is failed to achieve the trust between the end user and cloud service provider. Hence vendor uses this marvellous application should make trust. Weak trust relationship and lack of customer trust cause many problems during deployment of cloud services.

4. Study of Security Methods

In this section we are summarizing the different security methods those are employed in cloud computing and big data platforms in different ways.

In [1], authors Cheng-Kang Chu and Wen-Guey Tzeng proposed another security method of proxy re-encryption in which Alice temporarily delegating the task of decryption to the Bob through the third party proxy. As Alice giving the re-encryption key to proxy, proxy converts the ciphertext on behalf of Alice into the ciphertext on behalf of Bob. Two different identity based methods of proxy re-encryption are proposed by authors of this paper and shown security for standard model. The first method achieves the efficiency for ciphertext length as well as computation, whereas second module achieves selected ciphertext security. The problems those are mentioned in method of [2] are overcome by this method [1]. However, the problems with this method are that it does not providing the solutions for collusion resistance (CR), conditional share (CS) and anonymity.

In [2], author's proposed new identity based re-encryption security method. In this paper, author presented the novel framework which allowing the non-interactive and unidirectional proxy based re-encryption under the IBE systems. In this security method, existing proxy re-encryption method is extended to the identity based encryption (IBE) area in which data from the source is encrypted using the destination identity as public key. The advantage of this approach is it can be realized with new applications such as access control as well as attribute based delegation. However, this approach is still required to be efficient as does not addressing the problems like discovering the efficient constructions for methods of CCA-secure multi-use IBE-PRE. Also, discovering the efficient IBE-PRE method in standard model is open problem to address.

In [3], authors proposed new method to overcome the problems of method presented in [4]. This method is based method presented by Wang et al. previously for achieving the problem of WRO. The method called MUIBPRE (concrete multi-use unidirectional IBPRE in which security is shown for random oracle. However, the method of Wang et al. was having the drawback of not resisting against the collusion attack, problem of constructing MUIBPRE for standard model etc. This problem is overcome by this method presented in [3] where security framework is introduced the conversion from the NaHIBE (non-anonymous hierarchical identity based encryption) with the

CPA security to CCA-secure strongly as well as CR MUIBPRE. Clearly this method is showing advantage of achieving solutions against problems like WRO, CR and MU for efficient security framework. However, as our main research is on Big-Data platform security, there are two security problems are still open such as CS and anonymity.

In [4], authors presented the reviews of different ID based proxy re-encryption security methods called as IBPRE in which proxy with re-encryption key to transform ciphertext of one ID to another ciphertext of another ID. Such two cipher texts resulted into the same plaintext original message whereas the proxy does not able to get plaintext. As listed problems in existing methods by author, they mentioned that existing methods not fulfilling the complete requirements of security, therefore they proposed the new anonymous IBPRE (AIBPRE) method which is extension IBPRE method by including the security model and definition. In addition to this, both assumptions such as bilinear Diffie-Hellman (DBDH) and modified bilinear Diffie-Hellman (MDBDH) are used as decisive system for random oracle model. This approach shows the efficiency and security as compared to previous methods. In this method anonymity and CR is achieved along with security, but failed to provide solutions to CS, multiple ciphertext receiver update (MU) and without random oracle (WRO) problems.

In [5], JunJie Qiu, JungBok Jo and HoonJae Lee et al. introduced another variant of IBPRE based security method for cloud computing based applications. This method is based on CT07 framework introduced in [7]. The main goal of this method was to prevent the collusion of delegate and proxy. This method extends the approach proposed in [7] by overcoming its problems related to non-interactivity and uni-directionality. In this approach security is achieved by addition of secrete parameter as well as changing the re-encryption key and secrete key. The theoretical analysis of this method was showing the security against the CCA and collusion attack for standard model. The open research problem with this method is that CS and anonymity.

In [6], Xu a Wang, Xiaoyuan Yang et al. constructed the new security method based on identity based encryption method. The author's main contribution was use of concept of master key. This master key is embedded into the private key with goal of providing strong security against any kind of security attacks. In addition to this, some kind of randomness was introduced in private key in order to protect against the attacks attempted to extract sensitive information. It was shown by author that this new method is IND-sID-CPA secure for standard model based on the assumption of DBDH approach in bilinear groups. The advantage of this approach is that this IBPRE scheme which key of re-encryption is completely independent of private key of delegate's. And hence this IBPRE based method achieve the master secrete security. With this method limitation is of anonymity problem is not yet resolved.

In [8], authors Kaitai Liang, Willy Susilo proposed the new improved security framework especially for big data storage platforms. Above methods are showing the different advantages and limitations like anonymity problem is not resolved in most of methods. The client's anonymity is one of the major challenges of privacy in standard models and big data platforms and hence required to be addressed along with other security requirements. Therefore, authors of this paper first time proposed security approach in which along with other security requirements achieves the anonymity solution. In this approach privacy preserving ciphertext multi-sharing approach in order to overcome the problems associated with previous methods. This is hybrid approach in which anonymous method is combined with proxy re-encryption technique in which ciphertext is conditionally and securely shared many times by not leaking the ID of ciphertext senders or receivers as well as underlying message. The advantage of this approach is that it achieves MU, WRO, CR, anonymity, CS problems resolving. The limitation of this method is that as this is just theoretical proof of concepts, practical evaluation results not yet done and hence computation costs and communication costs may play important factor along with time required for data sharing.

In [9], authors recently presented new framework of security in order to provide the secure personal sensitive data sharing over big data platform with inclusion of tasks such as secure storage, secure usage, and secure data deliver as well as destruction over the semi trusted platform of big data. Author introduces the proxy re-encryption based method using heterogeneous ciphertext transformation as well as method of user process protection over Virtual machine monitor (VMM). This system provides the protection of end users sensitive data security efficiently and shares such data over big data platform securely. In addition to this, data owners are keeping the complete control on its own data under the healthy environment. The advantage of this method is that it achieves the security against different attacks such as collusion and CCA. Also achieves the solution of secure data destruction which was not with previous methods. The problem with this approach is that it does not supporting multi-user ciphertext approach like [8] method as well as this method not yet been practically evaluated.

5. Comparative Study

The methods which we discussed in above section are proposed for data security and privacy preserving in cloud computing systems. This section presenting the comparative study of all this methods in terms of various performance metrics like multiple ciphertext receiver update (MU), Conditional share (CS), Collusion resistance (CR), Anonymity, Results (RS), Secure Data Destruction (SDD) and efficiency. Table 2 showing the performance for total 8 methods those are recently presented.

Table 2: Comparative Analysis of Existing and Proposed Methods

Ref. No.	MU	CR	CS	Anonymity	SDD	Eff.	RS
[1]	Yes	No	No	No	No	No	No
[2]	Yes	No	No	No	No	Yes	No
[3]	Yes	Yes	No	No	No	Yes	Yes
[4]	No	Yes	No	Yes	No	No	Yes
[5]	Yes	Yes	No	No	No	Yes	No
[6]	Yes	Yes	Yes	No	No	Yes	No
[8]	Yes	Yes	Yes	Yes	No	No	No
[9]	Yes	Yes	Yes	No	Yes	Yes	No

From above table it is showing the recent methods showing the improvements in many factors such as MU, CR, CS, anonymity support, SDD and efficiency, however the most of methods seems to be not evaluated practically hence this is first limitation of existing methods. We listed total 8 methods in above table, out of which only 2 methods are practically evaluated. Another limitation is the recent method presented in [9] is does not supporting the anonymity as well as practically not evaluated.

6. Conclusion and Future Work

Goal of this paper is to present the study and analysis of existing cloud security methods by considering the big data systems as a part of it. This presents the survey over the cloud computing and its growth progress along with security surveys conducted by ICT. In this paper we also discussed the various approaches presented for cloud and big data security approaches along with their advantages and disadvantages. Finally, we presented the comparative study of all recent methods by considering the various security parameters. As this study is presented by considering future research works, in next work we will be presenting the novel framework for overcoming the current limitations of existing methods in cloud security.

References

[1] C.-K. Chu and W.-G. Tzeng. Identity-based proxy re-encryption without random oracles. In ISC '07, vol. 4779 of LNCS, pp. 189–202. Springer, 2007.
[2] M. Green and G. Ateniese. Identity-based proxy re-encryption. In ACNS '07, vol. 4512 of LNCS, pp. 288–306. Springer, 2007.
[3] J. Shao and Z. Cao. Multi-use unidirectional identity-based proxy re-encryption from hierarchical identity-based encryption. Inform. Sci.,2012. http://dx.doi.org/10.1016/j.ins.2012.04.013.
[4] J. Shao. Anonymous id-based proxy re-encryption. In ACISP, vol. 7372 of LNCS, pp. 364–375. Springer, 2012.
[5] JunJie Qiu, JungBok Jo and HoonJae Lee. Collusion-Resistant Identity-Based Proxy Re-Encryption Without Random Oracles. International Journal of Security and Its Applications Vol.9, No.9 (2015), pp.337-344
[6] Xu a Wang, Xiaoyuan Yang. New Identity Based Encryption and Its Proxy Re-encryption, The International Conference on Biomedical Engineering and Computer Science (ICBECS2010), May 3, 2012.
[7] Kalyani Shirudkar, Dilip Motwani. Big-Data Security. Volume 5, Issue 3, March 2015 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering.
[8] Kaitai Liang, Willy Susilo. Privacy-Preserving Ciphertext Multi-Sharing Control for Big Data Storage. IEEE Transactions on Information Forensics and Security, 1556-6013 (c) 2015 IEEE.
[9] Xinhua Dong, Ruixuan Li, Heng He, Wanwan Zhou, Zhengyuan Xue, and Hao Wu. Secure Sensitive Data Sharing on a Big Data Platform. TSINGHUA SCIENCE AND TECHNOLOGY ISSNl l1007-0214l l08/11l lpp72-80, Volume 20, Number 1, February 2015.
[10] S. Razick, R. Mocnik, L. F. Thomas, E. Ryeng, F. Drabløs, and P. Sætrom, the eGenVar data management system —Cataloguing and sharing sensitive data and metadata for the life sciences, Database, vol. 2014, p. bau027, 2014.
[11] Dikaiakos, M.D; Katsaros, D.; Mehra, P.; Pallis, G.; Vakali, A.; (2010), "Cloud Computing Distributed Internet Computing for IT and Scientific Research". Vol.13, pp 10, Sept.-Oct. 2009.
[12] Shuai Z; Shufen Z; Xuebin C; Xiuzhen H; (2010), "Cloud Computing Research and Development Trend", 2nd International conference on Future Networks, 2010. ICFN ' 10. pp 23, 22-24 Jan 2010.