A NOVEL IMPLEMENTATION OF ATTRIBUTE BASED ENCRYPTION FOR QUERY OPTIMIZATION IN BIG DATA USING MAPREDUCE

Dr. P. Vijaya Bharati[1], R. Ravi[2], N. Sowjanya Kumari[3]
[1]Associate Professor, [2]& [3]Assistant Professor
[1,2&,3] Department of Computer Science and Engineering,
[1,2,3] Vignan's Institute of Engineering for Women,
[1]pvijayabharati@gmail.com, [2]ravi4345@gmail.com, [3]sowjanyanalam5@gmail.com

Abstract

Data increased daily and has a significant role in every field, like industries, medical, etc. The data is captured, stored, and it is processed to retrieve the necessary data. Security and privacy play an essential role when critical data is shared among users in a distributed environment. These challenges are to be addressed. Mainly they are highly required during sharing and storing vast amounts of data. This paper presents a novel solution to secure the vast data with Attribute-based encryption (ABE), providing access control that prevents unauthorized user's access. Moreover, query optimization is provided in this paper to retrieve the required encrypted data from the big data quickly.

Keywords: Attribute-based encryption (ABE), Query Optimization, Mapreduce.

1. Introduction

Attribute-based encryption (ABE) [1, 2] was introduced in 2006 as a generalization of identity-based encryption (IBE) [3, 4]. ABE exists in two forms key-policy ABE (KP-ABE) and ciphertext-policy ABE (CP-ABE). In KP-ABE, the key master secret key and a master public key are generated, and each user is defined with a policy. The key is sent to the user, and a corresponding secret key is also generated with the user's defined policy. The sender needs to select some required attributes from the set of universal user attributes to encrypt any message and generates the ciphertext. The ciphertext can be decrypted by the receiver using only the matching secret key [47]. The difference between CP-ABE and KP-ABE is that each user has its attribute subset, and a ciphertext corresponds to a policy function f.

Several significant results are proposed to realize ABE in the last few years. These schemes have several types. One can be implemented to predicates computable by Boolean formulas [2, 6–11] (which are limited to log-depth computations). Another of these has made some critical progress [12–16], which can apply to sophisticated circuits. In 2013, based on the Learning with Errors (LWE) problem, Gorbunov et al. proposed a KP-ABE scheme [16] called GVW13, in which the predicate is the arbitrary polynomial-size circuits. So for an ABE system with a more number of users, and an efficient revocable mechanism is essential and vital.

The network's size has been increasing hugely, and data security in transmission became an important issue that needs to be considered [29]. The sender always wants to send the data more securely, ensuring that an eavesdropper does not read his message. To protect the data-in-transit, several encryption algorithms are regularly used in which the sender sends the encrypted data encrypting using the receiver public key. The receiver can recover the encrypted message into direct data only by the authorized user with their associated private key. The users share the key to communicate between two users where the sender sends the data, and the receiver receives it. However, in today's world, where high-speed internet is available, and several applications evolving, data is complex and stored in distributed environments [30]. The data storage may be on a local server or a remote server like an unknown cloud environment. In this type of environment, the data storage service provider should give users access to their data such that the other unauthorized

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

users should not be able to access that data publicly or store the data. This type of environment, as with the traditional encryption algorithms, is not sufficient. Instead, fine-grained access control of the data only with the legitimate users is required [31]. To do so, a ciphertext attribute-based encryption algorithm with query optimization and map-reduce is proposed in this paper.

2. Related Work

Data had rapid growth in the usage of the internet. A mechanism to provide trust and security is required to protect the sensitive and vital data that is to be stored in huge on data servers. Several encryption algorithms encrypt and protect the data, such as DES, AES, RC5, Diffie Hellman, blowfish, ElGamal, etc. These algorithms are robust but are breakable if the key is with anyone, either he is a legitimate user or an attacker [32-35]. Access control has become a crucial mechanism that enforces user access policies only to users who are authorized. A fine-grained access control methodology is being implemented where an access control policy is defined to access that data during encryption. The fundamental problem with existing algorithms is that they are prone to cryptanalysis since the same ciphertext is generated with the same key any number of times. If the key is lost, the encrypted data can easily be decrypted. There are other numerous types of attacks that can be performed on modern cryptosystems. Linear cryptanalysis, integral cryptanalysis, mod-n cryptanalysis, partitioning cryptanalysis, Traffic analysis, interpolation attack, brute force attack, Boomerang attack, timing attack, etc., are a few of the attacks. Initially, identify based encryption (IBE) is used, which encrypted and decrypted data using the users' identities, but the drawbacks of IBE is that the identities of the user are revealed.

To overcome their problem, ABE was proposed by Sahai and Waters in 2005. ABE provides both security and access control. It is a public-key encryption algorithm that allows users to encrypt and decrypt data using user's attributes. The secret key of the user and the ciphertext generation depend on attributes. The ciphertext can only be decrypted, only if the set of attributes of the user key matches the ciphertext attributes. An adversary that holds multiple keys should only be access data if at least one individual key grants access [37]. In this implementation of ABE, the user can decrypt the ciphertext only if the required secret key that matches the attributes used in encrypting the plain text is provided. On the other side, even adversaries can get some data parts. However, they cannot be understood as the relation with the whole big data is different. Hence, the proposed method will encrypt only some parts of the data randomly. Generally, the big data path is at the level of l bytes that can be processed and stored easily. It can also be transmitted at ease.

Query optimization plays a vital role in retrieving data. The amount of time spent on the implementation of an optimized query plan and the quality are trade-offs. They should be balanced, and there are different ways to balance them. Generally, the query plan execution is implemented in a tree where the nodes are arranged in the tree structure. The intermediate results are retrieved from the bottom of the tree to the top of the tree, which finally gives the optimized result. As this implementation takes more execution time, the proposed work does query optimization in a distributed manner, explained in the next section.

3. Proposed System

Security plays a significant role in the data stored in a remote server or a cloud environment at a remote place. To store and perform computation in big data, the Hadoop platform is chosen, a distributed environment where big data can be stored efficiently. Additional security measures are required in a distributed system than a centralized system due to numerous users using different data with distributed control. In the distributed environment, the intruders are categorized as passive attackers who listen to the message communicated on the network and active attackers who listen to the message transmitted and modify and fabricate the message on the network. To ensure security on the network, data should be prevented from attacks by the intruder. In distributed environments, the security of data in transit and the security of data at rest need to be adopted.

To ensure that only authenticated users can access the data, access control measures need to be adopted

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

. To protect the data, encryption techniques are implemented. Data encryption can be done in two ways. The first is that the data is encrypted by the users and stored in an encrypted format. The users fetch the encrypted data and then decrypt it. Key management is the responsibility of the user himself. The second way is that the storage provider does the encryption. For security purposes, the users store and retrieve the data without realizing that data is stored in an encrypted form.

In this proposed system, a cipher–policy attribute-based encryption is used. ABE has four algorithms, as shown in figure 3.1, and they are:

1. Setup
2. Key generation
3. Encrypt
4. Decrypt

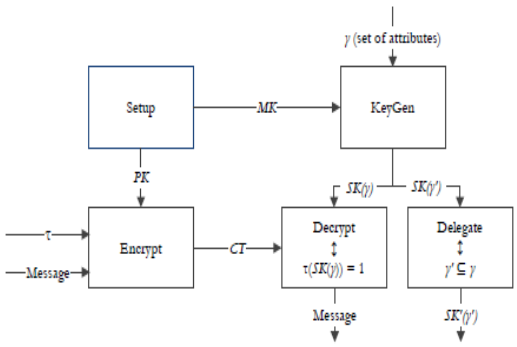

Figure 1: CPABE Construction

1. Setup (P):

The setup algorithm takes security parameters P as the input and generates the master public key MK.

2. Key generation (MK, U):

This algorithm has taken the master public key MK and set of users attributes U as input, and generates a secret key SK that satisfy the access policy defined with the user attributes.

3. Encrypt (MK, M, A):

This algorithm has taken as input master public key MK defined with access structure A over the attribute F ∈ U, the set of attributes. The algorithm encrypts the message M and produces the ciphertext CT.

4. Decrypt (SK, CT, F):

This algorithm takes the secret key SK, ciphertext CT over the attribute F as input and decrypts the ciphertext to plain message M if and only the attribute F given as input to decrypt the message satisfies the access structure; otherwise, the decrypted message is garbage.

To secure the data stored, CPABE is implemented, and often, data is outsourced in an encrypted format, which protects data from illegal access. Nevertheless, the problem is with data searching, which should be carried out accurately. Hence, query optimization is applied to facilitate the execution of the user's exact query matching.

3.1 Distributed Query Processing

In the distributed environment, a query's processing is optimized at the name node and data node. The data is fragmented into chunks, and replicas of the chunks are stored across multiple data nodes to enhance the data's robustness. The name node has the complete path information and the information about the data stored in the chunks. The user requests are processed at the name node, and mapping is made to the relevant data node. The steps involved are as follows:

1. The name node validates the user, and the query is checked, translated, and optimized at the name node itself.
2. In the next step, the name node optimization generates an execution plan and handles the query processing to the job tracker.
3. The job tracker divides the query into multiple task handles, and the tasks are assigned to the task trackers, who in turn reply to the job tracker on completion of the job.
4. The data required has been retrieved from the data nodes, and queries are optimized. Finally, the query results are merged and sent by the task trackers to the job tracker and, in turn, to the name node.
5. The name node is finally responsible for sending the result to the queried users.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

There is a possibility of adversaries getting some chunks of data but they cannot understand the data as the data chunks are been stored in an encrypted format.

4. Analysis and Comparison

4.1 Efficiency of the proposed scheme

If the data is transmitted locally, there are no issues with the data's size, but if the data from the user to be transmitted to a remote server, it is a difficult task. In the proposed scheme, the divided data chunks are stored on different storage servers, as long as a certain amount of redundant data backup strategy is chosen. Even if some of the storage servers fail, the data stored will not be affected as the data's replicas are maintained. In the proposed model, it is easy to prevent unauthorized access when only the data's authorized user wants to encrypt the data. It is practically impossible to encrypt a massive amount of data stored directly in real-time applications. This model is easy to share the data with other users by distributing the secret attribute and defining other users' access policies.

4.2 Security of the proposed scheme

If the proposed method's vulnerability is computed, let x be an adversary who wants to access the data illegally. However, the proposed method defines an access policy such that only an authorized user can access the data, and if the unauthorized user still tries to decrypt the data, the data retrieved is some garbage. In the proposed model, data is separated into n chunks, and these n chunks will be stored as a minimum of three replicas at m different storage servers. On considering some situations, they assumed that the adversary x knows the secret key of the chunks of data in probability p such that $0<p<1$. The probability that the adversary x achieves the access policy on the data is pm, and when the probability p is less, and the number of servers is more in number, the probability pm will be very low.

4.3 Comparison with other schemes

As data is increasing and users are also increased, initially, there are particular concerns. Despite the challenges raised, it is reluctant to store the vast data in servers and maintain them such that the data is made available on need by numerous users. Later, the data complications with the security and privacy of the data played a significant role. In work [38], a survey on different security risks that focus on different security issues is concerned. The traditional encryption schemes for protecting data are mentioned, and different novel schemes [39-40] have been proposed. The Table 1 represents comparison of different ABE Schemes.

Table 1: Comparison of different ABE Schemes

Criteria ABE KP-ABE CP-ABE ABE with non monotonic Criteria ABE KP-ABE CP-ABE ABE with non Criteria	ABE	KP - ABE	CP-ABE	ABE with Query Optimization
Fine Grained Access Control	-	Satisfied	Satisfied	Satisfied
Scalability	-	-	-	Satisfied
User Accountability	-	-	Satisfied	Satisfied
User Revocation	-	Satisfied	Satisfied	Satisfied

Comparing the proposed model with a traditional scheme is figured in figure 2.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

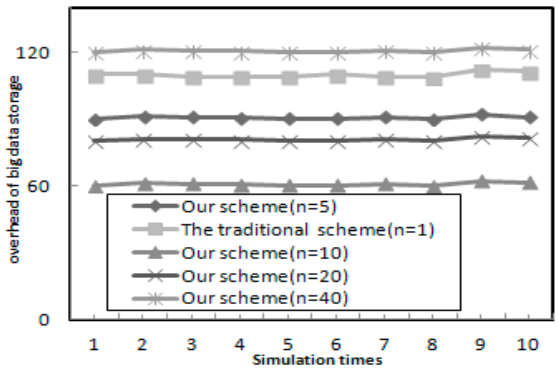

Figure 2: Implementation of two schemes on big data storage

The figure 2 represents that using the traditional mechanisms, and even if the workload is less, it takes time for query processing, but for the proposed scheme, even if the workload increases almost simultaneously, the queries are processed, showing the best results. The figure 3 compares CP-ABE with proposed scheme ABE with Query optimization and shows good results. The accuracy of retrieving the queried data of a particular user is more compared to existing ABE implementation in Big data.

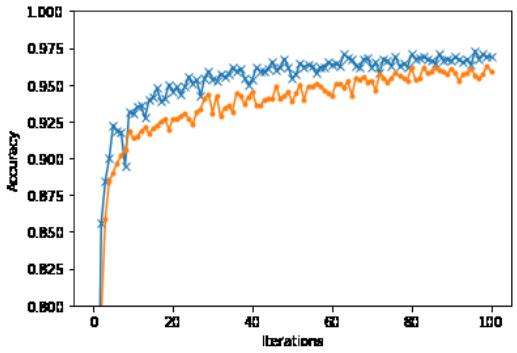

Figure 3: Comparison the accuracy of CP-ABE and CP-ABE with Query Optimization

5. Conclusion

As the data's size is enormous, data encryption is costly, which is of very much concern. Finally, this paper presents a novel solution to tell the importance of the study to protect users' security who wants to access the stored data at a remote place. The proposed techniques improve the performance in preserving access to the individual user. Even though practitioners' adopting these techniques is a difficult task, its provision for high security makes it to be adapted.

ABE with query optimization supports only one privilege level, so it is not appropriate for quick access management. It also helps to free from the procedure and to challenge to perform complicated operations. The experimental results show that the ABE with query optimization is versatile with user accountability, collusion resistant, and user revocation even with big data. However, it's still slow in some devices due to an integrated mathematical process operation structure. Using specific algorithms in the access policy, the attributes are accustomed to generating a public key to encipher the information and a secret key consisting of user attributes to rewrite the information. Further, it can be enhanced with Hierarchical ABE, where an access policy to each user is employed, providing more security restricting the user's access.

References

6.1 Journal Article

[1] Applebaum, B., Beimel, A., Farràs, O., Nir, O., Peter, N., 2019. “Secret-sharing schemes for general and uniform access structures, in: Annual International Conference on the Theory and Applications of Cryptographic Techniques, Springer. pp. 441-471.

6.2 Journal Article

[2] Camenisch, J.L., Lehmann, A., Neven, G., 2019. Production of cryptographic signatures. US patent App. 16/374,197.

6.3 Journal Article

[3] Bethencourt J, Sahai A, Waters B, Ciphertext-policy attribute-based encryption 2007 IEEE symposium on security and privacy (SP’07), IEEE (2007), pp. 321-334

6.4 Journal Article

[4] Boneh, D., Boyen, X., Goh, E.J., 2005. Hierarchical identity based encryption with

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

constant size ciphertext, in: Annual International Conference on the Theory and Applications of Cryptographic Techniques, Springer. pp. 440-456.

6.5 Journal Article

[5]
Bozovic V, Socek D, Steinwandt R, Villanyi V.I., Multi-authority attribute-based encryption with honest-but-curious central authority, International Journal of Computer Mathematics, 89 (2012), pp. 268-283

6.6 Journal Article

[6] Chen, Y., Li, W., Gao, F., Yin, W., Liang, K., Zhang, H., Wen, Q., 2019, Efficient attribute-based data sharing scheme with hidden access structures, The Computer Journal.

6.7 Journal Article

[7] Chen Y, Martínez J.F., Castillejo P, López LA bilinear map pairing based authentication scheme for smart grid communications: Pauth IEEE Access, 7 (2019), pp. 22633-22643

6.8 Journal Article

[8] Cui H, Deng R.H., Lai J, Yi X, Nepal SAn efficient and expressive ciphertext-policy attribute-based encryption scheme with partially hidden access structures, revisited Computer Networks, 133 (2018), pp. 157-165

6.9 Journal Article

[9]
Gavriloaie R, Nejdl W, Olmedilla D, Seamons K.E. , Winslett MNo registration needed: How to use declarative policies and negotiation to access sensitive resources on the semantic web European Semantic Web Symposium, Springer (2004), pp. 342-356

6.10 Journal Article

[10] Gharahi M, Khazaei SOptimal linear secret sharing schemes for graph access structures on six participants Theoretical Computer Science, 771 (2019), pp. 1-8

6.11 Journal Article

[11] Goyal, V., Pandey, O., Sahai, A., Waters, B., 2006. Attribute-based encryption for fine-grained access control of encrypted data, in: Proceedings of the 13th ACM conference on Computer and communications security, Acm. pp. 89-98.

6.12 Journal Article

[12] Han, Y., Lu, W., Yang, X., 2013. Attribute-based signcryption scheme with non-monotonic access structure, in: 2013 5th International Conference on Intelligent Networking and Collaborative Systems, IEEE. pp. 796-802.

6.13 Journal Article

[13] Harney H, Colgrove A, McDaniel P Principles of policy in secure groups, NDSS, Citeseer (2001)

6.14 Journal Article

[14] Kang, M.H., Park, J.S., Froscher, J.N., 2001. Access control mechanisms for inter-organizational workflow, in: Proceedings of the sixth ACM symposium on Access control models and technologies, ACM. pp. 66-74.

6.15 Journal Article

[15] Koppula, V., Waters, B., 2019. Realizing chosen ciphertext security generically in attribute-based encryption and predicate encryption, in: Annual International Cryptology Conference, Springer. pp. 671-700.

6.16 Journal Article

[16] Lee C.C., Chung P.S., Hwang M.S.A survey on attribute-based encryption schemes of access control in cloud environments IJ Network Security, 15 (2013), pp. 231-240

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

6.17 Journal Article

[17] Li Y, Zhang P, Wang BAn improved ciphertext-policy attribute-based encryption scheme in power cloud access control Applied Sciences, 8 (2018), p. 1836

6.18 Journal Article

[18] Nishide, T., Yoneyama, K., Ohta, K., 2008. Attribute-based encryption with partially hidden encryptor-specified access structures, in: International conference on applied cryptography and network security, Springer. pp. 111-129.

6.19 Journal Article

[19] Ostrovsky, R., Sahai, A., Waters, B., 2007. Attribute-based encryption with non-monotonic access structures, in: Proceedings of the 14th ACM conference on Computer and communications security, ACM. pp. 195-203.

6.20 Journal Article

[20] Reddy, N.P.K., Reddy, E.K., 2018. Fine grained access control using attribute-based encryption (abe) technique in cloud computing.

6.21 Journal Article

[21] Sahai, A., Waters, B., 2005. Fuzzy identity-based encryption, in: Annual International Conference on the Theory and Applications of Cryptographic Techniques, Springer. pp. 457-473.

6.22 Journal Article

[22] Shoup, V., 1997. Lower bounds for discrete logarithms and related problems, in: International Conference on the Theory and Applications of Cryptographic Techniques, Springer. pp. 256-266.

6.23 Journal Article

[23] Wang J, Java realization for ciphertext-policy attribute-based encryption, Computer Science College of Shandong University (2012)

6.24 Journal Article

[24]
Xue L, Yu Y, Li Y, Au M.H., Du X, Yang BEfficie nt attribute-based encryption with attribute revocation for assured data deletion Information Sciences, 479 (2019), pp. 640-650

6.25 Journal Article

[25]
Yamada S, Attrapadung N, Hanaoka G, Kunihiro N A framework and compact constructions for non-monotonic attribute-based encryption. International Workshop on Public Key Cryptography, Springer (2014), pp. 275-292

6.26 Journal Article

[26] Zhang K, Li H, Ma J, Liu XEfficient large-universe multi-authority ciphertext-policy attribute-based encryption with white-box traceability Science China Information Sciences, 61 (2018), p. 032102

6.27 Journal Article

[27]
Zhao Y, Fan P, Cai H, Qin Z, Xiong HAttribute-based encryption with non-monotonic access structures supporting fine-grained attribute revocation in m-healthcare IJ Network Security, 19 (2017), pp. 1044-1052.

6.28 Journal Article

[28] Zhong H, Zhu W, Xu Y, Cui JMulti-authority attribute-based encryption access control scheme with policy hidden for cloud storage Soft Computing, 22 (2018), pp. 243-251

6.29 Journal Article

[29] D. Kusnetzky. What is "Big Data?" [Online]. Available: http://blogs.zdnet.com/virtualization/?p=1708.

6.30 Journal Article

[30] Shmueli, Erez, et al. "Database encryption: an overview of contemporary challenges and design considerations." ACM SIGMOD Record 38.3 (2010): 29-34.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

6.31 Journal Article

[31] Spoorthy V, Mamatha M, Kumar B S., "A Survey on Data Storage and Security in Cloud Computing," 2014.

6.32 Journal Article

[32] S. Subashini，V. Kavitha，"A survey on security issues in service delivery models of cloud computing"，Journal of Network and Computer Applications, Vol 34, Issue 1, January 2011, pp. 1–11.

6.33 Journal Article

[33] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, pp. 612–613, November 1979. [Online]. Available: http://doi.acm.org/10.1145/359168.359176.

6.34 Journal Article

[34] Y. Singh, F. Kandah, and W. Zhang, "A secured cost-effective multicloud storage in cloud computing," in Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on, April 2011, pp. 619 –624.

6.35 Journal Article

[35] M. G. Jaatun, G. Zhao, A. Vasilakos, A . A. Nyre, S. Alapnes, and Y. Tang, "The design of a redundant array of independent net-storages for improved confidentiality in cloud computing," Journal of Cloud Computing: Advances, Systems and Applications, vol. 1, no. 1, p. 13, 2012.

6.36 Journal Article

[36] J. Spillner, G. Bombach, S. Matthischke, J. Muller, R. Tzschichholz, and A. Schill, "Information dispersion over redundant arrays of optimal cloud storage for desktop users," in Utility and Cloud Computing, 2011 Fourth IEEE International Conference on, Dec. 2011, pp. 1 –8.

6.37 Journal Article

[37] Qian Wang; Cong Wang; Kui Ren; Wenjing Lou; Jin Li. Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing. IEEE Transactions on Parallel and Distributed Systems, vol. 22(5), 2011, pp.847-859.

6.38 Journal Article

[38] Boneh D and Franklin M, "Identity-based encryption from the Weil pairing" in Advances in Cryptology, CRYPTO 2001, Lecture Notes in Computer Science, vol. 2139, pp. 213-229,2001.

6.39 Journal Article

[39] Trummer, Immanuel, Koch, Christoph, "Multi-Objective Parametric Query Optimization", 2015, pp: 221–232.

6.40 Journal Article

[40] Ioannidis, Yannis, Raymond, Shim, Kyuseok, Sellis, Timos, "Parametric Query Optimization", 1997, pp:132–151

6.41 Journal Article

[41] Trummer, Immanuel, Koch, Christoph," Approximation Schemes for Many-Objective Query Optimization, 2014, pp:1299–1310.

6.42 Journal Article

[42] Selinger, P. G.,Astrahan, M. M., Chamberlin, D. D., Lorie, R. A., Price, T. G., "Access Path Selection in a Relational Database Management System", 1979, Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data, pp. 23–34.

6.43 Journal Article

[43] Chaudhuri, Surajit ,"An Overview of Query Optimization in Relational Systems", Proceedings of the ACM Symposium on Principles of Database Systems, 1998, pp. 34–43.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

6.44 Journal Article

[44] Ioannidis, Yannis, "Query optimization". ACM Computing Surveys, March 1996, pp: 121–123.

6.45 Journal Article

[45] Kumaran, Thamil, et al. "An Impact of Implementing Various Cryptographic Techniques Efficiently in a Public Centric Cloud." International Journal of Science, Engineering and Computer Technology, vol. 4, no. 4, Indian Association of Health, Research and Welfare, Apr. 2014, p. 83.

6.46 Journal Article

[46] Cloud Services : Security and Compliance with Data .... https://blogs.cisco.com/datacenter/cloud-services-security-and-compliance-with-data-sovereignty-law

6.47 Journal Article

[47] Lee, Kwangsu. “A Generic Construction for Revocable Identity-Based Encryption with Subset Difference Methods.” PLoS One, vol. 15, no. 9, Public Library of Science, Sept. 2020, p. e0239053.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731