International Journal of Safety and Security Engineering
Vol. 10, No. 3, June, 2020, pp. 417-422
Journal homepage: http://iieta.org/journals/ijsse

A Crypto Scheme Using Data Obfuscation of Entity Detection and Replacement for Private Cloud

Yakobu Dasari*, Hemanth Kumar Kalluri, Venkatesulu Dondeti

Department of CSE, Vignan's Foundation for Science, Technology and Research, Vadlamudi, AP 522213, India

Corresponding Author Email: dy_cse@vignan.ac.in

https://doi.org/10.18280/ijsse.100315

Received: 10 January 2020
Accepted: 3 April 2020

Keywords:
cloud computing, confidentiality, obfuscation, cryptography

ABSTRACT

Cloud has been rising, renown, and extremely demanding innovation now a day. Cloud has wide ubiquity with its advanced features, like web access, more stockpiling, easy setup, programmed refreshes, low cost, and resource provisioning on a rent basis. Disregarding many advantages, security is viewed as increasingly significant and drew the consideration of numerous researchers. The information storage is drastically increasing, and there are many occasions that cloud doesn't ensure that data/information that has been placed in the cloud is secured from unauthorized access. Many experts are attempting to guarantee data security in the cloud, yet tragically they don't give satisfactory results. Hence we attempted to propose an effective crypto-scheme with obfuscation and cryptography for unstructured information. The scheme attempts to safeguard the secrecy of information at two phases. In the first phase, it obfuscates the file by supplanting the keywords (obfuscation), and at the subsequent phase, the obfuscated file is encoded by using the conventional RSA (Rivest Shamir Adleman) encryption algorithm for high security. Investigation results show that the proposed mechanism yields great outcomes.

1. INTRODUCTION

Advancement of Technology from over the years has come a long way. One of such was the idea of cloud computing. It helps to solve the problems of traditional data storage systems. Its cost is very less when compared to other sources and easy to use. The client pays as he uses cloud assets [1]. Majorly it is profitable for those who have tiny enterprises and self-experts who expand their applications without having large investments. It economically eases as he pays what he uses. There are three modules where this technology mainly works. The primary is service providing person [2], who is responsible for letting information be kept in the cloud. A secondary module is a consumer who can have permission to access the essential services from the cloud. And the last is an interface of the web by which users can keep resources in the cloud and even can retrieve the services that are required [3]. These services are mainly classified as software as a service (SaaS), infrastructure as a service (IaaS), Application as a service (AaaS), and Platform as a service (PaaS), Communication as a service, Network as a service and Database as a service [4]. In the cloud, these services will be deployed through private, public, and hybrid models [5]. Cloud has very significant features that differentiate it from other technologies. They are pay-as-you-go [5], auto-scaling [6], flexibility [7], elasticity. According to the pay-as-you-go process of billing is done based on the utilization of the amount of the resources. This is playing an essential role in the upward of this technology.

Regardless of having such more prominent features, it thinks of security issues [7]. The security issues are associated with data [8], network, infrastructure, privacy, assaults, and so on [9, 10]. The significant research work concentrated on giving security to information in the cloud from being accessed by unauthorized clients [11]. This is known as data/information security [12]. Data that is put away in the cloud ought to be kept private as per the necessities of the cloud client. Cloud still not ensured data security [13, 14]. Information that is placed/shared through cloud commonly incorporates client transaction data, client inclinations, and patient's health records, etc. Subsequently, applications in cloud research ordinarily require compelling strategies/mechanisms to protect the security of the cloud client's sensitive data [15]. These strategies/mechanisms should work viably by appropriately handling, testing, and examining the huge amount of data. Real-time information, which has the most sensitivity, is generally attractive for checking and approving the viability and appropriateness of these strategies and mechanisms.

There is a need for strategies and mechanisms that are viable and effective to guarantee the protection of such client data before it is made open [16, 17]. A few methods that scramble or supplant the delicate information in reports in a way that doesn't uncover the sensitive or the private data. These methods guarantee that when files are shared, the ill-conceived client couldn't catch the client's delicate data yet preserves the readability of information [18]. Other methods incorporate cryptography, which converts the information into some other form that can't be read and understand by anybody except if it is transformed into its original form [19, 20]. In the principal case, however, it doesn't give genuine data, there is as yet a chance of information breach since it permits the readability. In the subsequent case, however, it doesn't empower the readability, there is a likelihood that outsider can decrypt the

file and use it. Thus, mechanisms ought to be planned with the end goal that no secret data is uncovered to the ill-conceived client, yet the file ought to be in a way that other analytical process can be performed on the file [14, 21, 22].

In our work, we designed and implemented an effective crypto-scheme for preserving the privacy of unstructured data such as company policies [23]. Our mechanism protects the data in two phases. In the first phase, it replaces the first data with sham data by detecting and supplanting the keywords in the file using Natural Language Processing (NLP) methods [24, 25] and permits the readability of the file to such an extent that it is conceivable to reproduce the original file. Retaining the readability of the file has two advantages. (I) when an outsider (ill-conceived client) get to gain access to file, he can't conclude which information is original and which is supplanted, (ii) since enough data accessible, we can perform analytical and processing Application on file. In the second phase, it transforms the delicate data into an unreadable form using a conventional RSA algorithm.

The overview of the paper is as follows. Section-II shows the related works done on ensuring data security in cloud computing. The working of proposed scheme has been described in section-III. Section-IV provides experimental results of proposed scheme with sample data. Section-V ends with the conclusion.

2. LITERATURE SURVEY

By using the signal method, Hashemzade and Maroosi [18] introduced hybrid signal and encryption obfuscation strategy for changing structure program of tree and graph into a star-like structure and hides flow control graph of the problem. The main disadvantage of signal strategy is that it has many numbers of call and return instructions. By observing this it's been recommended to add a dispatcher for the program, which transforms the signal program to that of the flow graph. This dispatcher was added because it prevents unauthorized users and keeps us secure. Rather than this, another application has been prescribed to measure difficulty and robustness. When compared with the results method, which is proposed presents an advantage over the existing strategies.

Data Obfuscation of text data using entity detection and replacement techniques was introduced by S.V Balakrishnan, which ensures the confidentiality of users' sensitive data. The proposed technique identifies the key entities which are used to coordinate the configuration parameters and tags them for obfuscation.

The tagged entities are then replaced with dummy entities. A hash table is used to retrieve the dummy entities. The hash table can also be used to give a turnaround time of scrambling strategy by which novel data can be restored to an obfuscated file.

Iwaya et al. [26] showcased a mechanism called privacy-preserving ontology-based obfuscation, whose main purpose was to obfuscate information related to health either for first and foremost or for later use. On account of first utilization reducing information of individuals describes that a stager will not get information than needed by a combination of the semantic level fitting. The later use, the mechanism that is being proposed, can have a limited loss of information, with the final goal topmost level of use is kept, by the time where information is kept secret to described requirements.

Arul Oli and Arockiam [27] proposed a methodology called "Confidentiality technique using data obfuscation" for improving the security of information that is stored in the public cloud. There is numerical data in the cloud, and this methodology helps in protecting this data. Encrypting the incompatible numerical information by the user by involving confusion, this methodology is fit and flexible. To encrypt and decrypt, two keys are utilized in this algorithm. And the two keys that are used are values of integers. By using these two keys, non-transparency of data that is numerical is feasible through ARO_Obfus_CT for preserving data in the cloud.

3. PROPOSED METHODOLOGY

The proposed mechanism targets making sure about the unstructured information secure, for example, patients' health records and any organizational policies that have been shared through private cloud. It ensures the protection of delicate data. It ensures the security of information at two phases: At first phase it replaces the first data with sham data by detecting and supplanting the keywords in the file using NLP methods [24, 25] and permits the readability of the file to such an extent that it is conceivable to reproduce the original file. Retaining the readability of the file has two advantages. (I) when an outsider (ill-conceived client) gain access to file, he can't conclude which information is original and which is supplanted, (ii) since enough data is accessible, we can perform analytical and processing Application on file. In the second phase, it transforms the delicate data into the unreadable form using a conventional RSA algorithm. The structure of our proposed scheme is shown in below Figure 1.

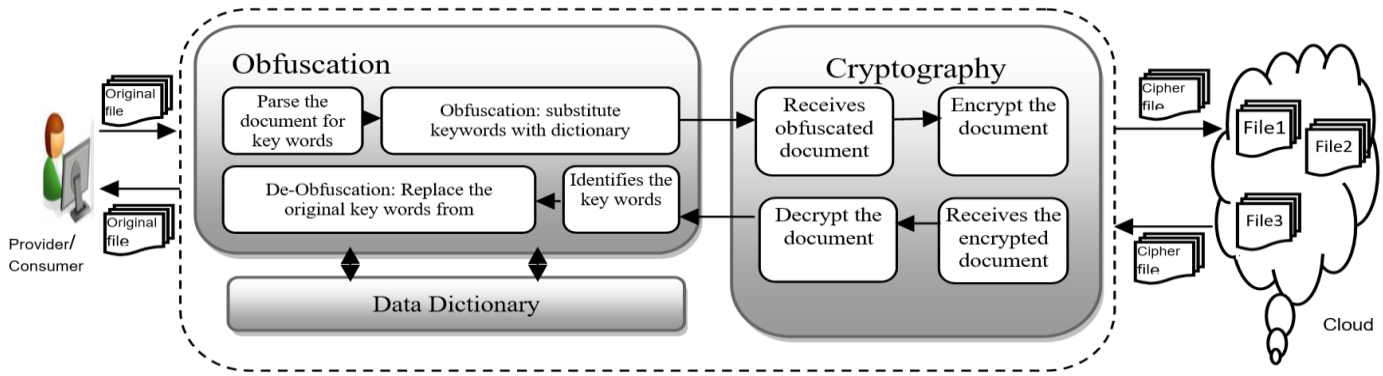

Figure 1. Structure of the proposed scheme

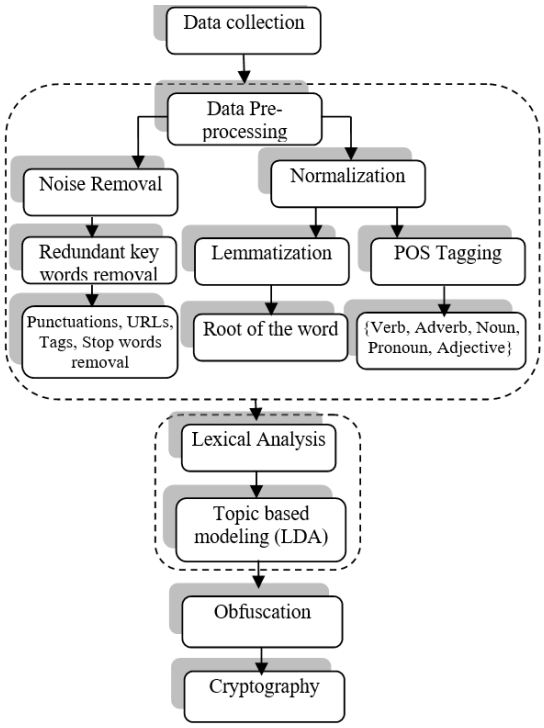

Figure 2. Workflow of mechanism

The proposed algorithm has three modules, Obfuscator and Cryptanalysis module.

Obfuscator: This module plays out the obscurity and de-obscurity operations on an input file that contains unstructured information. At whatever point a client needs to store his delicate data in the cloud, the obfuscator module scramble the data in the file and sends it to the crypt-analysis module. Furthermore, de-scramble the file, at whatever point client needs to use it. This module endeavors to holds the readability of the file to such an extent that it is conceivable to reproduce the original file [28, 29].

Cryptanalysis: Usually cryptanalysis module transforms the information into a form that can't be read and understand by other users who are not authorized. It takes the obfuscated document and encrypts it by using a public key encryption algorithm RSA, which can be implemented practically and yields better results than the rest of the algorithms [30, 31]. The workflow algorithms are shown in Figure 2.

Initially, the sample data is collected for obfuscation. To extract the keywords from the document we perform the following steps.

Data pre-processing: Data pre-processing includes two steps they are (i) Noise removal and (ii) Normalization.

Noise removal: Various data pre-processing techniques have been applied to the data for noise removal. Techniques have been used to remove the redundant keywords, punctuations, URLs, tags, and stop words, etc. [25]. Stop words refer to a large number of prepositions, which are to be filtered out before processing the data.

Normalization: Normalization is a process of converting a list of words to a more uniform sequence. This is useful in preparing text for later processing. Techniques like converting the text into lower case, parts of speech (POS) tagging, lemmatization has been applied on the data for smooth processing. Lemmatization and POS tagging are the special cases of normalization [30].

Lemmatization: Lemmatization is a more advanced technique that works based on the root of the word.

POS tagging: It is the way toward labeling a word in a corpus to a corresponding part of a speech tag, in accordance with its context and definition. POS Tagging is a significant step to comprehend the meaning of any sentence or to extricate a relationship and fabricate a knowledge graph. Python NLP bundle is utilized for actualizing the procedures.

Lexical analysis: It includes recognizing and investigating the structure of words. The vocabulary of a language implies the assortment of words and expressions in a language. The lexical examination is separating the entire lump of content into passages, sentences, and words [31].

Topic based modelling: Topic vocabulary has been built using Latent Dirichlet Allocation (LDA) model. Based on the number of topics constructed from LDA, a set of seed words will be extracted for building the Topic-Based Modeling.

A set of words for each topic can be extracted by using the below parameters.

$$P(\theta_{1:M}, Z_{1:M}, \beta_{1:k} / D; \alpha_{1:M}, \eta_{1:k})$$

where, $'\theta'$ denote the distribution of topics, $'z'$ a number of topics for each report $'\beta'$ signify the dispersion of words for every topic. Given $'D'$ denotes the document $'\alpha'$ and $'\eta'$ denotes the parameter vectors for document and topic, respectively.

But this can't be calculated nicely because this entity is intractable. To solve this problem, variational inference technique has been used to minimize the Kullback-Leibler divergence (KL divergence) between the approximation and true posterior as an optimization problem using the below Eq. (1).

$$\gamma^*, \emptyset^*, \lambda^* = argmin_{(\gamma,\emptyset,\lambda)} D(q(\theta, Z, \beta/\gamma, \phi, \lambda)// p(\theta, z, \beta/D;\ \alpha, \eta) \quad (1)$$

where, γ, ϕ and λ denotes the free variational parameters we surmised θ, z and β with, separately. Here $D(q||p)$ denotes the KL divergence among q and p. What's more, by changing γ, ϕ, and λ, we get distinctive q appropriations having various distances from the genuine posterior p. We will likely discover the γ^*, ϕ^*, and λ^* that limit the KL divergence between the surmised q and the genuine posterior p.

Obfuscation: Data obfuscation is a form of data veiling where information is intentionally mixed/supplanted with sham information to prevent unauthorized access to sensitive materials [18]. This form of scrambling/replacement results in unintelligible or confusing data. Data obfuscation at this point scrambles the data in a manner that retains readability. All the keywords which are extracted in the previous step would be replaced with keywords in the custom-built dictionary by applying the "FlashText algorithm". Figure 3 shows an example of obfuscating the record-2, in which original information is replaced with dummy information, which is colored in yellow.

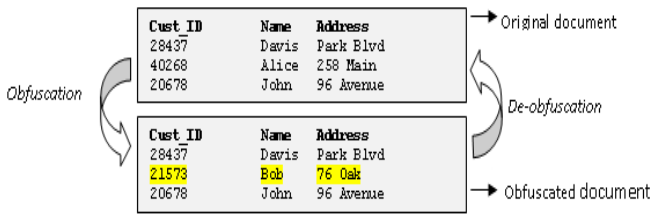

Figure 3. Data obfuscation example

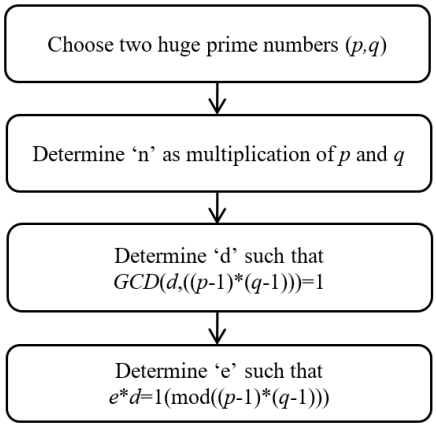

Figure 4. Process of determining *n, d,* and *e* values

Cryptography: Is a process of keeping an information secret. i.e. it transforms the information into a form that can't be read and understand by other users who are not authorized. It has four concerns, confidentiality, integrity, non-repudiation, and authentication. In our work, we use cryptography techniques to ensure the confidentiality information stored in the cloud.

At this level, the obfuscated document will be encrypted by using a public key encryption algorithm RSA, which can be implemented practically and yields better results than the rest of the algorithms. The working procedure of RSA is given below.

The encryption and decryption process are done using Eq. (2) and Eq. (4) respectively.

$$C = M\text{^}e \; mod \; n \tag{2}$$

$$M = C\text{^}d \; mod \; n \tag{3}$$

where, $C \; and \; M$ are the ciphertext and Plaintext.

Here the value of 'n', should be known to the cloud service provider as well as the cloud service consumer. And service provider knows the value of 'e' and only authorized consumer knows the value of 'd'. Here public key is $\{ e, n \}$ and private key is $\{d , n \}$.

And, to determine the values of e, d, n we have to follow the steps below.

Step-1: Choose two numbers p and q which are huge primes.

Step-2: Calculate 'n' as the multiplication of p and q.

Step-3: Choose some large integer randomly, d, such that$GCD(d, ((p-1)*(q-1))) = 1$.

Step-4: Find the value of 'e' such that $e * d = 1(mod((p-1)*(q-1)))$.

Based on these values and following the RSA procedure, the input gets converted to ciphertext. The process of determining *n, d, e* values is shown in Figure 4.

4. EXPERIMENTAL RESULTS AND ITS ANALYSIS

Our mechanism is executed on the AWS cloud environment, and execution of the equivalent is checked with existing mechanisms as far as the time is taken for both obfuscation/encryption and decryption/de-obfuscation. The results show that the security level of the proposed mechanism is higher than the existing mechanisms. The existing technique works on structured numerical data, whereas the proposed scheme works effectively on unstructured data.

The efficiency of the proposed scheme lies in exact keywords extraction based on semantics, substituting keywords with dummy keywords (obfuscation), and substitution of dummy keywords with original keywords (de-obfuscation). The no. of topics selected, no. of key words extracted, time taken for key words extraction and hit ratio of key words extraction from various documents of different sizes is shown in Table 1.

Table 1. Hit ratio of key words extraction from documents of different sizes

File size (Kb)	No. of of topics	No. of key words	Time taken for keywords extraction (MilliSec)	% Keyword hit
1Kb	2	10	21	100%
10Kb	20	100	48	100%
50Kb	100	215	63	100%
100Kb	200	450	82	100%
200Kb	400	923	97	100%
400Kb	800	1765	221	100%
600Kb	1200	2534	325	100%
800Kb	1600	3324	487	100%
1000kb	2000	4312	561	100%

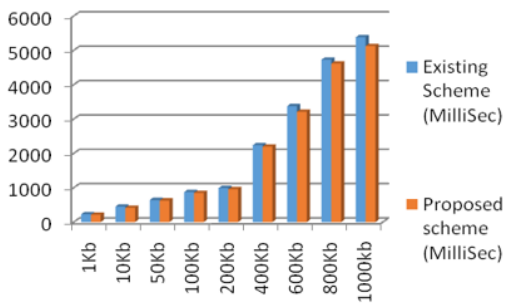

Figure 5. Encryption & obfuscation processing of proposed and existing mechanisms

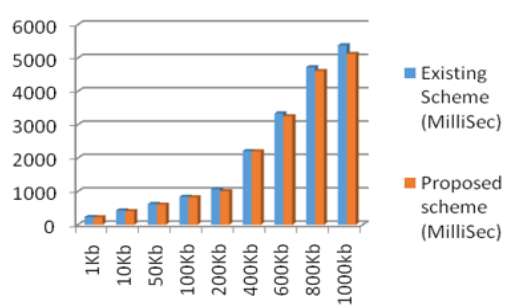

Figure 6. Decryption and de-obfuscation processing time of existing and proposed mechanism

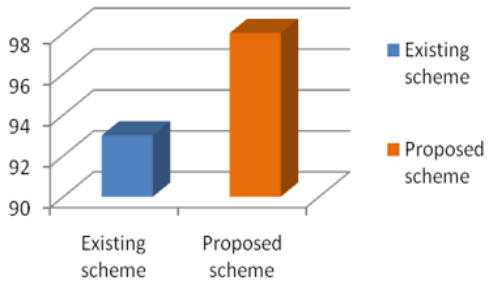

Figure 7. Comparison between security levels of existing and proposed schemes

Table 2. Processing time of existing and proposed techniques for encryption and obfuscation

File Size (Kb)	Existing Scheme (MilliSec)	Proposed scheme (MilliSec)
1Kb	225	211
10Kb	452	412
50Kb	641	627
100Kb	873	848
200Kb	989	956
400Kb	2239	2198
600Kb	3374	3211
800Kb	4726	4613
1000kb	5382	5123

Table 3. Decryption and de-obfuscation processing time of existing and proposed mechanism

File Size (Kb)	Existing Scheme (MilliSec)	Proposed scheme (MilliSec)
1Kb	232	224
10Kb	434	412
50Kb	623	602
100Kb	840	824
200Kb	1063	1003
400Kb	2212	2201
600Kb	3339	3256
800Kb	4726	4613
1000kb	5382	5123

Table 4. Level of security offered by existing and proposed mechanism

Crypto Techniques	Security level
Existing scheme	93
Proposed scheme	98

Figure 5 shows that the proposed mechanism takes less effort for both obfuscation and encryption than existing approaches.

Figure 6 shows that the proposed mechanism takes less effort for both obfuscation and encryption than existing mechanisms.

Figure 7 shows that the security level offered by our mechanism is a lot higher than the existing mechanism.

Proposed scheme takes less time for processing obfuscation & encryption when compared with existing schemes, which is shown in Table 2.

Table 3 shows the processing time taken by proposed scheme for decryption & de-obfuscation, which is found to be less when compared to existing schemes.

The level of security offered by proposed scheme is very high when compared to existing schemes, that is shown in below Table 4.

5. CONCLUSION

In this work, we made an attempt to propose an effective crypto-scheme with obfuscation and cryptography to secure user's sensitive information in the cloud environment. Proposed scheme saves the secrecy of information in two phases effectively. The proposed scheme experimented in the AWS cloud environment and performance of the same is checked with existing strategies regarding the accurate key words extraction and time taken for both obfuscation/de-obfuscation and encryption/decryption. The experimental result shows that the proposed scheme works effectively in keywords extraction, obfuscation, de-obfuscation, encryption and decryption. It takes less time for performing obfuscation/de-obfuscation and encryption/decryption. The test results also show that the security level of the proposed scheme is higher than the existing strategies. The robustness of proposed scheme is also at satisfactory level with phase margin of 30o and gain margin of 3db.

REFERENCES

[1] Khorshed, M.T., Ali, A.S., Wasimi, S.A. (2012). A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. Future Generation Computer Systems, 28(6): 833-851. http://dx.doi.org/10.1016/j.future.2012.01.006

[2] Pathan, M., Vecchiola, C., Buyya, R. (2008). Load and proximity aware request-redirection for dynamic load distribution in peering CDNs. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", pp. 62-81. http://dx.doi.org/10.1007/978-3-540-88871-0_8

[3] Halabi, T., Bellaiche, M. (2018). A broker-based framework for standardization and management of Cloud Security-SLAs. Computers & Security, 75: 59-71. http://dx.doi.org/10.1016/j.cose.2018.01.019

[4] Chowdary, E.D., Yakobu, D. (2016). Cloud of Things (CoT) integration challenges. 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, pp. 1-5. https://doi.org/10.1109/ICCIC.2016.7919553

[5] Chauhan, S.S., Pilli, E.S., Joshi, R.C., Singh, G., Govil, M.C. (2019). Brokering in interconnected cloud computing environments: A survey. Journal of Parallel and Distributed Computing, 133: 193-209. http://dx.doi.org/10.1016/j.jpdc.2018.08.001

[6] Ferrer, A.J. (2016). Inter-cloud research: Vision for 2020. Procedia Computer Science, 97: 140-143. http://dx.doi.org/10.1016/j.procs.2016.08.292

[7] Hussain, S.A., Fatima, M., Saeed, A., Raza, I., Shahzad, R.K. (2017). Multilevel classification of security concerns in cloud computing. Applied Computing and Informatics, 13(1): 57-65. https://doi.org/10.1016/j.aci.2016.03.001

[8] Sun, Y., Zhang, J., Xiong, Y., Zhu, G. (2014). Data security and privacy in cloud computing. International Journal of Distributed Sensor Networks, 10(7): 190903.

https://doi.org/10.1155/2014/190903
[9] Vurukonda, N., Rao, B.T. (2016). A study on data storage security issues in cloud computing. Procedia Computer Science, 92: 128-135. http://dx.doi.org/10.1016/j.procs.2016.07.335
[10] Wang, R. (2017). Research on data security technology based on cloud storage. Procedia Engineering, 174: 1340-1355. http://dx.doi.org/10.1016/j.proeng.2017.01.286
[11] Ab Rahman, N.H., Choo, K.K.R. (2015). A survey of information security incident handling in the cloud. Computers & Security, 49: 45-69. https://doi.org/10.1016/j.cose.2014.11.006
[12] Saurabh, S., Young-Sik, J., Jong, H.P. (2016). A survey on cloud computing security: Issues, threats and solution. Journal of Network and Computer Applications, 75(2016): 200-222. http://dx.doi.org/10.1016/j.jnca.2016.09.002
[13] Rohit, B., Sanyal, S. (2012). Survey on security issues in cloud computing and associated mitigation techniques. International Journal of Computer Applications, 47(18): 0975-0888. http://dx.doi.org/10.5120/7292-0578
[14] Usman, M., Jan, M.A., He, X. (2017). Cryptography-based secure data storage and sharing using HEVC and public clouds. Information Sciences, 387: 90-102. http://dx.doi.org/10.1016/j.ins.2016.08.059
[15] Vijayalakshmi, M., Yakobu, D., Veeraiah, D., Rao, N.G. (2016). Automatic healing of services in cloud computing environment. 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, pp. 740-745. http://dx.doi.org/10.1109/ICACCCT.2016.7831738
[16] Garikapati, G., Yakobu, D., Nitta, G., Amudhavel, J. (2017). An analysis of cloud data security issues and mechanisms. International Journal of Pure and Applied Mathematics, 116(6): 141-147.
[17] Yakobu, D., Shanmugam, M. (2019). A hybrid secure mechanism for effective storage confidentiality to ensure data integrity and privacy in public cloud. JARDCS, 11(7): 330-339.
[18] Hashemzade, B., Maroosi, A. (2018). Hybrid obfuscation using signals and encryption. Journal of Computer Networks and Communications, 2018: 6873807. http://dx.doi.org/10.1155/2018/6873807
[19] Jorstad, N.D., Landgrave, T.S. (1997). Cryptographic algorithm metrics. 20th National Information Systems Security Conference, pp. 1-38.
[20] Ghosh, A., Saha, A. (2013). A numerical method based encryption algorithm with steganography. Computer Science & Information Technology, 3: 149-157. https://doi.org/10.5121/csit.2013.3214
[21] Horváth, M., Buttyán, L. (2015). The birth of cryptographic obfuscation-A survey. Cryptology ePrint Archive, Report 2015/412.
[22] Murthy, T.S., Gopalan, N.P., Yakobu, D. (2019). An Efficient un-realization algorithm for privacy preserving decision tree learning using McDiarmid's bound. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(4S2): 499-502.
[23] Monikandan, S., Arockiam, L. (2015). Confidentiality technique to enhance security of data in public cloud storage using data obfuscation. Indian Journal of Science and Technology, 8(4): 1-10. 10.17485/ijst/2015/v8i24/80032
[24] Hirschberg, J., Manning, C.D. (2015). Advances in natural language processing. Science, 349(6245): 261-266. http://dx.doi.org/10.1126/science.aaa8685
[25] Lewinski, N.A., McInnes, B.T. (2015). Using natural language processing techniques to inform research on nanotechnology. Beilstein Journal of Nanotechnology, 6(1): 1439-1449. http://dx.doi.org/10.3762/bjnano.6.149
[26] Iwaya, L.H., Giunchiglia, F., Martucci, L. A., Hume, A., Fischer-Hübner, S., Chenu-Abente, R. (201). Ontology-based obfuscation and anonymisation for privacy. IFIP International Summer School on Privacy and Identity Management, pp. 343-358. https://doi.org/10.1007/978-3-319-41763-9
[27] Arul Oli, S., Arockiam, L. (2015). Confidentiality technique for data stored in public cloud storage. International Journal of Engineering and Technical Research, V5(2): 169-174. http://dx.doi.org/10.17577/IJERTV5IS020028
[28] Hohenberger, S., Rothblum, G.N., Shelat, A., Vaikuntanathan, V. (2011). Securely obfuscating re-encryption. Journal of Cryptology, 24(4): 694-719. http://dx.doi.org/10.1007/s00145-010-9077-7
[29] Ravikumar, G.K., Rabi, B.J., Manjunath, T.N., Hegadi, R.S., Archana, R.A. (2011). Design of data masking architecture and analysis of data masking techniques for testing. International Journal of Engineering Science and Technology, 3(6): 5150-5159.
[30] Sun, S., Luo, C., Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. Information Fusion, 36: 10-25. http://dx.doi.org/10.1016/j.inffus.2016.10.004
[31] Sevenster, M., Buurman, J., Liu, P., Peters, J.F., Chang, P.J. (2015). Natural language processing techniques for extracting and categorizing finding measurements in narrative radiology reports. Applied Clinical Informatics, 6(3): 600-610. http://dx.doi.org/10.4338/ACI-2014-11-RA-0110