ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 3 Issue 1 January 2023 / Page No: 15-20
Paper Id: JETA-V3I1P103 / Doi: 10.56472/25832646/JETA-V3I1P103

Review Article

Privacy Preservation Techniques in Cloud Computing

Himani Saini1, Gopal Singh2
[1,2]*Department of Computer Science and Application, Maharshi Dayanand University, Rohtak (Haryana), India*

Received Date: 03 January 2023 Revised Date: 17 January 2023 Accepted Date: 28 January 2023

Abstract: *Cloud Computing helps customers to take benefit from the variety of services, such ason-demand self-service, universal network connectivity, risk transfer, a usage-based payment andlocation-independent resource sharing. The privacy of cloud data is a critical concern in the cloudcomputing environment that necessitates unique attention. K-anonymity is a useful concept for preserving privacy in cloud computing. The concept of generalization is most commonly used because of its little information loss. However, such algorithms are often computationally expensive, they struggle to perform well when dealing with massive amounts of data. This articleexamines the security risks and concerns that cloud consumers and service providers confront. This research study identifies the research gaps and proposes a method to address the problem.*

Keywords: *Cloud Computing, k-anonymity, Privacy Preservation, Security, Swarm Intelligence*

I. INTRODUCTION

Cloud computing is a crucial part of todays advanced computing technologies. Computation's ideas, computing and architectures have changed and strengthened through the years. Cloud Computer is a computing environment that is rapidly reallocating resources as the next step in thedevelopment and deployment of large numbers of distributed applications [1].

The common service computation system used by the majority of cloud computation providers isinspiring for clients whose digital asset requirements change over time. This is a censorious layoutparameter in modern data hub and cloud computing applications. These high energy costs and highcarbon emissions are largely due to the electricity and energy produced by software machinery and the linked heating unit [2].

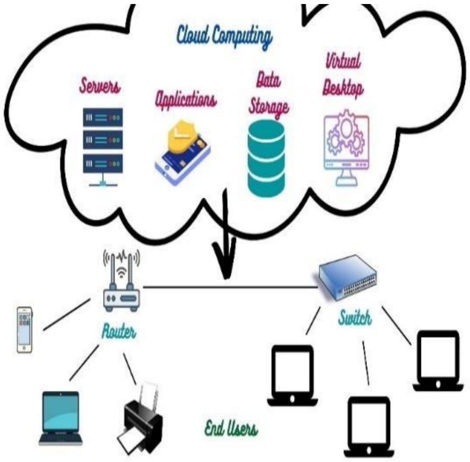


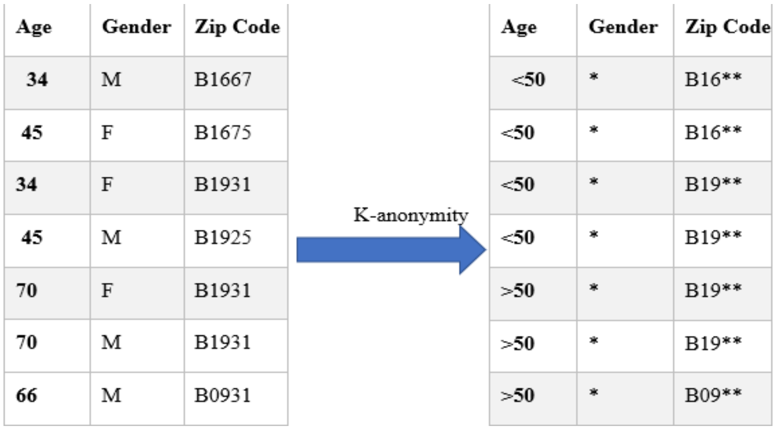
Figure 1: Cloud Computing Architecture

The benefits of using cloud computing are numerous as long as the privacy and security issues areidentified and appropriately mitigated. People are now recognizing the benefits of the Cloud. Thecloud offers various services like [3].

a) Scalability: To handle changing workloads, cloud infrastructure scales on demand.
b) Storage: Depending on security requirements, users can opt for public, private, or hybrid storage options.
c) Accessibility: Cloud-based applications can be accessed from almost any internet-connecteddevice.
d) Security: Due to networked backups, data loss is not a concern when hardware fails.
e) Speed: Users may easily bring their apps to market by developing on the cloud.
f) Every day, data is generated as a result of every technological action, and the data size is rising exponentially.
g) This makes maintaining privacy more difficult. To solve this difficulty, the k-anonymity methodology is developed.

A. K-Anonymity

It's a method for keeping social media data private. When sharing information, statistics or any company information on social media, there may be a link between them that makes data sensitiveto the third party. The degree of a specific group in the network is (k-1) elements if the data has the k-anonymity property. As a result, the information on the network is secure. The primary keyis changed with dummy identifiers or extra noise nodes during the release of records, so that intruders in the network cannot identify them [4]. Even after the dummy identifier has been added,some traits result in trying to identify data breach, also known as a quasi-identifier. Some table attributes, such as birth date, gender, and zip code traits, can reveal personal information when combined with publicly available data, such as the voting list table, resulting in data vulnerability.



Age	Gender	Zip Code
34	M	B1667
45	F	B1675
34	F	B1931
45	M	B1925
70	F	B1931
70	M	B1931
66	M	B0931

K-anonymity

Age	Gender	Zip Code
<50	*	B16**
<50	*	B16**
<50	*	B19**
<50	*	B19**
>50	*	B19**
>50	*	B19**
>50	*	B09**

Figure 2: Anonymized data after both Generalization and Suppression

This type of privacy breach is avoided with K–anonymity, which confirms the release of an individual's record if there are (k-1) distinct individuals with related records which are not separated from the previous [5]. A social network may be used by an intruder to attack the victim (G). A neighbor= (u) who is associated with a social network, i.e., u V (G). Neighbor (u) appears k times in G', where G' is theanonymization of G. Then, with 1/k reliability, the neighbor (u) can be easily identified in G'.

To protect the privacy of the vertices, the re-identification confidence threshold value in the k- anonymity model is set lower [6]. Vertex= u ϵ V (G) U = k-anonymous in anonymization (G') only if there are at least (k-1) verticesand the neighborhood G'(A(u), A(v1), A(vk-1) is isomorphic. Only G' will be k-anonymous if each vertex is k-anonymous.

B. Issues in Privacy Preservation

Privacy preservation aims to secure the data in order to reduce the security threats and other aspectsof the data owner. There are various issues which gives raise to privacy concern especially when the data owner keeps his/her data where there is no centralized control. The end users may not be willing to share their geolocation details, such as latitude and longitude, with third parties, as mentioned by Pillai et al. in their quantitative analysis study. The study provides insightful details about end-user behavior.

There are two concerns in privacy mechanism namely the distribution of degree of the data owner and identification of sensitive attribute in the data-list. If a cloud network is viewed as a graph G (V, E) in which V represents the total number of vertices and E represents the total number of edges., it is said to be safe if the node stands an equivalent Degree Distribution (DD).

$$DD = \frac{\text{T}otal\ number\ of\ data\ owner\ having\ same\ degree}{\text{T}otal\ number\ of\ data\ owner\ in\ the\ current\ cluster}$$

The degree of a data owner is the total sum of in-degree and out-degree.

In degree: The connection which comes inward to the node is called the in-degree and the connection that goes outward is termed as out degree as shown in Figure 3.

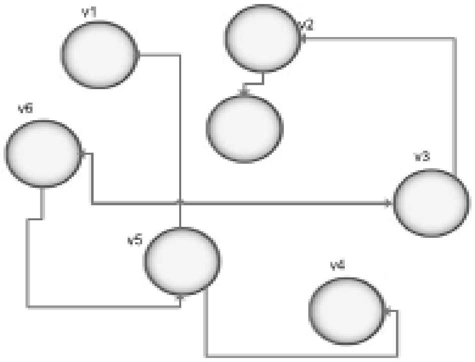


Figure 3: Network Diagram

In this small scale plotted network, there are 6 vertices and 6 edges. As for example, V2 has an in-degree of 1 as it gets a connection from V3 and has an out-degree 1 as it provides the connection an anonymous data owner. The risk of transfer or data storage gains a privacy issue here. Hence in order to neutralize the graph structure, the graph is to be divided in such a manner that k-1 datanodes get the same degree value so that the intruder does not bifurcate in any term.

As information cannot be measured directly, so the loss can be evaluated by calculating the loss inthe gain of the data. The gain can be evaluated by transforming the data into frequency domain and then calculating the noise present in the data. Both the aspects are responsible for raise and fall in Information Loss. Some recent submitted articles have shown maturity in utilizing the modern-day distance formulation as mentioned in related work [7].

II. RELATED WORK

Cloud computing has become the integral part of the daily technology drivers and even in the commercial usage; common people have also started analysing and using it. With the increased usage, privacy related concerns have also gained a raise time by time. This section illustrates the related work in the same contrast.

Chang et al. (2017) have provided a comparative analysis to know different factors influencing user's trust on Face book and LinkedIn social sites. Data has been collected by organizing interviews in which the participants are subject domain experts. Based on the answers an empirical study has been performed to find relationship between different influencing factors like privacy concern, risk, trust and many more. Effort suspense, social influence, and superficial risk are determined as the main influencing factors on the trust of social networking Sites (SNS) [8].

Fei et al. (2017) proposed a two-tier security scheme based on the principle of anonymity while increasing privacy costs. In this article, a cost-sharing strategy is also proposed. To optimize the users' personal information sharing, first divide the region into small groups by maximizing the maximum entropy. Also, discuss how to determine which LBS user generates k-1 dummytrajectories and receives payments from the others by using the cost sharing problem. After that, they proposed a cost-sharing mechanism for keeping LBS private. The result demonstrates that, for the same cost, our approach can provide a stronger level of privacy [9].

Ma et al. (2018) to de-anonymize social networks, a random forest method was proposed. Becauseonline social networks share user data with researchers, advertisers, and developers, user privacy is at risk. These data are anonymized by discarding private information such as the user's identification name and personal address; however, users' personal information is still not adequately protected. The proposed research divides large sparse social networks into several tiny subgraphs using the spectral partition technique. The proposed method uses a random forest classifier to classify candidate node pairs from the anonymous and auxiliary networks as matchedpair. The outcomes show an increase in efficiency and accuracy[10].

Rao et al. (2018) suggested new privacy preservation techniques named FakeIt for social media data sharing. Today different sites provide GPS tracking services by using current user location via browsing, online applications, and Wi-Fi devices, which makes it easy for service providers toprovide user location services to access private user data and to access user details. To ensure thesecurity of the user's new FakeIt technique for privacy preservation has been proposed. The author showed that the technology proposed works around security, confidentiality in order to satisfy the requirements of privacy and people must decide on the information before sharing. If the present location shared by the user is very sensitive, the user must choose to share the wrong position on the basis of the current location with the service provider [11].

Sarah et al. (2018) suggested an anonymizing technique named Isomorphism technique for this research area, that prevents the privacy of individuals in the weighted social network. An isomorphism (R, S) technique is an anonymizing method used to produce an R1 candidate subgraph for each and every subgraph. An (R, S) – isomorphism depends on a variety of techniques to make each subgraph of R-1 similar subgraph as an example of weighted maximum frequency subgraphs, graph density, bi-clustering methods, and weighted protective measure have been discussed [12].

Aghasian et al. (2018) Suggested a privacy conservation technique and a violation of social network data such as Facebook, Instagram, Twitter, etc. Because the privacy risks are increasing, users must hide the relevant details from others. To minimize the risk of disclosure of secret information, they proposed a new technique that safeguards privacy. They gave users the idea of what kind of information they should share with other people. The proposed technique uses the Bernstein polynomial theorem to understand what information affects the privacy of the user. Afterwards, an anonymized data of people connected to someone on social media sites is appliedto a new model [13].

Krishnan and Chen (2019) have presented a cloud-based framework to recognized tweets from fake news using Machine Learning (ML) approach. Main components of the framework have been deployed as web services that can be merged or used in other applications. Using, this model, user interface to web has been designed to analyse the tweet credibility though web browser [14].

Wang et al. (2019) Propose the new concept of social media crowd computing invitation and rewarding privacy security (PPIR). The author shows the structure of PPIR. Finally, to formalize its system model and security model and to develop a concrete PPIR scheme, based on the bilinearpairings and group-oriented encryption technique. The random model proved that the proposed PPIR system was protected. The performance was measured by the connectivity and calculation costs. The safety and productivity study shows that our specific PPIR is safe and effective [15].

III. ANALYSIS OF EXISTING WORK

The existing work and its limitations are discussed in this section. We have also proposed the appropriate solutions to overcome such gaps.

Table 1: Existing Techniques their Limitations and Proposed Solution

licationYear	Title of the Paper	Author's	iques/ Toolused	Research Gaps	Proposed Solution
2018	Cloud Computing Fault Tolerance Aware Scheduling Technique using Dynamic ClusteringAlgorithm	Latiff etal [16]	Dynamic Clustering LeagueChampionship Algorithm (DCLCA) SchedulingApproach	Have not Testedthe System in Real Cloud Environment	Optimized System Performance by combining LCA with other appropriate Smart Scheduling Techniques
2018	Community Discovery in Online Social Networks withProtected Privacy	Zheng etal [17]	Enhanced Community detection approach with Privacy Preservation Algorithms	Apart from being from the same community, other factors could have been addressed	Future implementation of Social network community detection where social connections and profiles are delicate

2019	Privacy PreservationBig Data Publicationon Cloud using Mondrian Anonymization Technique and Deep Neural Network	Andrew et al [18]	Mondrian Based Anonymization Approach and Deep Neural Networks	Security issues increases in caseof Big Data	Q-learning could have been another option, if big data isconcerned
2019	Enhancing cloud service reliability through Dynamic Replication and DataMigration Mechanism	Kaur et al[19	Wireless Personal Communication	Modern day Algorithms were found to be missing such as Swarm Intelligence and other Machine Learning Architecture	Implementation of Swarm intelligence Algorithm improvesData Migration
2020	An Application of Crypto Cloud Computing in SocialNetworks by Co-operative Game Theory	Ergun etal [20]	Crypto Cloud Based Method Incorporation toGame Theory	Efficiency and Reliability was checked using only the AmazonWeb Services Data	Stochastic, Fuzzy or Grey Uncertainty canbe applied
2020	Smart Government Policy Monitoring and Control Using Social Networkingand Cloud Computing	Singh etal [21]	Proposed an Integrated Schemaby combining Cloud Computing with SNS	Bot Detection wasnot used and only used Twitter Data	Fog Computing can Overcome Such typeof Limitations

IV. RESULTS AND DISCUSSION

The existing framework like Hadoop Map Reduce, k-anonymity, which is used for large data processing faces big data challenges for optimizing large data that is growing day by day. In orderto alleviate the difficulty of performance, various kinds of performance assessment tools have been proposed and developed to accurately determine the performance of big data at runtime. The TABLE-1 shows the different problems and the limitations faced by using the existing Privacy Preservation Techniques and also discuss the various Solutions which can help to Overcome such limitations. The main objective of this paper is mainly the formation of clusters for the enhancement of k-anonymity and use of various similarity measures like Jaccard Similarity, Tanimato Co-efficient for the evaluation of K (number of nodes) value. The recommended solutions are practical and will assist future effort. In the near Future, we will work on the implementation of Swarm Intelligence Algorithm for the reduction of noisy nodes distribution.

V. CONCLUSION AND FUTURE WORK

Usage of ML and Artificial Intelligence (AI) is being observed in recent times for cloud computingwhich is missing for the privacy concerns. The problem of this research work is to enhance the privacy mechanism in contrast to two things that are Enhancement in DD and Enhancement in theidentification of sensitive attribute through distance measures. In this Paper, we discussed about Various Privacy Preservation Problems in Cloud Computing Environment. The table lists the various methods that have been researched, as well as the different proposed solutions. Different approaches to resolve the challenges that limit privacy preservation are also examined. We have suggested the various solutions to these problems.

VI. REFERENCES

[1] P. Krishnadoss and P. Jacob, “OCSA: Task scheduling algorithm in cloud computing environment,” Int. J. Intell. Eng. Syst., vol. 11, no. 3, pp. 271–279, 2018, doi: 10.22266/ijies2018.0630.29.

[2] S. A. Ali, M. Affan, and M. Alam, “A study of efficient energy management techniques for cloud computing environment,” Proc. 9th Int. Conf. Cloud Comput. Data Sci. Eng. Conflu. 2019, pp. 13–18, 2019, doi: 10.1109/CONFLUENCE.2019.8776977.

[3] J. Wang, Y. Zhao, S. Jiang, and J. Le, “Providing privacy preserving in Cloud computing,” 3rd Int. Conf. Hum. Syst. Interact. HSI’2010 - Conf. Proc., pp. 472–475, 2010, doi: 10.1109/HSI.2010.5514526.

[4] S. Zhang, X. Li, Z. Tan, T. Peng, and G. Wang, “A caching and spatial [Formula presented]-anonymity drivenprivacy enhancement scheme in continuous location-based services,” Futur. Gener. Comput. Syst., vol. 94, pp. 40–50, 2019, doi: 10.1016/j.future.2018.10.053.

[5] A. Campan and T. M. Truta, “Data and structural K-anonymity in social networks,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 5456 LNCS, no. February, pp. 33– 54, 2009, doi: 10.1007/978-3-642-01718-6_4.

[6] L. Sweeney, “Achieving k-anonymity privacy protection using generalization and suppression,” Int. J. Uncertainty, Fuzziness Knowlege-Based Syst., vol. 10, no. 5, pp. 571–588, 2002, doi: 10.1142/S021848850200165X.

[7] K. Arava and S. Lingamgunta, “Adaptive k-Anonymity Approach for Privacy Preserving in Cloud,” Arab. J. Sci. Eng., vol. 45, no. 4, pp. 2425–2432, 2020, doi: 10.1007/s13369-019-03999-0.

[8] S. E. Chang, A. Y. Liu, and W. C. Shen, “User trust in social networking services: A comparison of Facebook and LinkedIn,” Comput. Human Behav., vol. 69, pp. 207–217, 2017, doi: 10.1016/j.chb.2016.12.013.

[9] F. Fei, S. Li, H. Dai, C. Hu, W. Dou, and Q. Ni, “A K-Anonymity Based Schema for Location Privacy Preservation,” IEEE Trans. Sustain. Comput., vol. 4, no. 2, pp. 156–167, 2019, doi: 10.1109/TSUSC.2017.2733018.

[10] J. Ma et al., “De-Anonymizing Social Networks with Random Forest Classifier,” IEEE Access, vol. 6, no. c, pp. 10139– 10150, 2017, doi: 10.1109/ACCESS.2017.2756904.

[11] J. Rao, R. Pattewar, and R. Chhallani, “A Privacy-Preserving Approach to Secure Location-Based Data,” pp. 47–55.

[12] S. Al-Kharji, Y. Tian, and M. Al-Rodhaan, “A Novel (K, X)-isomorphism Method for Protecting Privacy in Weighted social Network,” 21st Saudi Comput. Soc. Natl. Comput. Conf. NCC 2018, pp. 1–6, 2018, doi: 10.1109/NCG.2018.8593107.

[13] E. Aghasian, S. Garg, and J. Montgomery, “A privacy-enhanced friending approach for users on multiple online social networks,” Computers, vol. 7, no. 3, 2018, doi: 10.3390/computers7030042.

[14] S. Krishnan and M. Chen, “Cloud-based system for fake tweet identification,” 2019 IFIP/IEEE Symp. Integr. Netw. Serv. Manag. IM 2019, pp. 720–721, 2019.

[15] H. Wang, D. He, and J. Yu, “Privacy-preserving incentive and rewarding scheme for crowd computing in socialmedia,” Inf. Sci. (Ny)., vol. 470, pp. 15–27, 2019, doi: 10.1016/j.ins.2018.07.016.

[16] C. zhi Gao, Q. Cheng, X. Li, and S. bing Xia, “Cloud-assisted privacy-preserving profile-matching scheme under multiple keys in mobile social network,” Cluster Comput., vol. 22, pp. 1655–1663, 2019, doi: 10.1007/s10586-017-1649-y.

[17] X. Zheng, Z. Cai, G. Luo, L. Tian, and X. Bai, “Privacy-preserved community discovery in online social networks,”Futur. Gener. Comput. Syst., vol. 93, pp. 1002–1009, 2019, doi: 10.1016/j.future.2018.04.020.

[18] S. M. Abdulhamid, M. S. Abd Latiff, S. H. H. Madni, and M. Abdullahi, “Fault tolerance aware scheduling technique for cloud computing environment using dynamic clustering algorithm,” Neural Comput. Appl., vol. 29, no. 1, pp. 279–293, 2018, doi: 10.1007/s00521-016-2448-8.

[19] J. Andrew, J. Karthikeyan, and J. Jebastin, “Privacy Preserving Big Data Publication on Cloud Using Mondrian Anonymization Techniques and Deep Neural Networks,” 2019 5th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2019, pp. 722–727, 2019, doi: 10.1109/ICACCS.2019.8728384.

[20] S. Ergün, B. B. Kirlar, S. Z. Alparslan Gok, and G. W. Weber, “An Application Of Crypto Cloud Computing In Social Networks By Cooperative Game Theory,” J. Ind. Manag. Optim., vol. 16, no. 4, pp. 1927–1941, 2020, doi: 10.3934/jimo.2019036.

[21] P. Singh, Y. K. Dwivedi, K. S. Kahlon, R. S. Sawhney, A. A. Alalwan, and N. P. Rana, “Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing,” Inf. Syst. Front., vol. 22, no. 2, pp. 315–337, 2020, doi: 10.1007/s10796-019-09916-y.

[22] Sanjaikanth E Vadakkethil Somanathan Pillai. (2021). Balancing Precision and Privacy: Harnessing Location-Based Services in Healthcare Delivery. International Journal on Recent and Innovation Trends in Computing and Communication, 9(12), 50–56. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10963

[23] Hu, Wen-Chen, Sanjaikanth E. Vadakkethil Somanathan Pillai, and Abdelrahman Ahmed ElSaid. "Mobile Health Text Misinformation Identification Using Mobile Data Mining," International Journal of Mobile Devices, Wearable Technology, and Flexible Electronics (IJMDWTFE) 12, no.1: 1-14. http://doi.org/10.4018/IJMDWTFE.311433

[24] Chanthati, Sasibhushan Roa. (2021). A segmented approach to encouragement of entrepreneurship using data science. World Journal of Advanced Engineering Technology and Science. https://doi.org/10.30574/wjaets.2024.12.2.0330. [Link]

[25] Naga Ramesh Palakurti, 2022. "AI Applications in Food Safety and Quality Control" ESP Journal of Engineering & Technology Advancements 2(3): 48-61.

[26] Ayyalasomayajula, Madan Mohan Tito, Srikrishna Ayyalasomayajula, and Sailaja Ayyalasomayajula. "Efficient Dental X-Ray Bone Loss Classification: Ensemble Learning With Fine-Tuned VIT-G/14 And Coatnet-7 For Detecting Localized Vs. Generalized Depleted Alveolar Bone." Educational Administration: Theory and Practice 28.02 (2022).

[27] Bhat, V. Gojanur, and R. Hegde. 2015. 4G protocol and architecture for BYOD over Cloud Computing. In Communications and Signal Processing (ICCSP), 2015 International Conference on. 0308-0313. Google Scholar. [Link]

[28] Aparna Bhat, Rajeshwari Hegde, “Comprehensive Study of Renewable Energy Resources and Present Scenario in India,” 2015 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, TN, India, 2015. [Link]

[29] Piyush Ranjan, 2022.”Fundamentals Of Digital Transformation In Financial Services: Key Drivers and Strategies”, International Journal of Core Engineering & Management, Volume 7, Issue 3, PP 41-50, [Link]