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 as on-demand self-service, universal network connectivity, risk transfer, a usage-based payment and location-independent resource sharing. The privacy of cloud data is a critical concern in the cloud computing 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 article examines 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 today’s 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 the development and deployment of large numbers of distributed applications [1].

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

The benefits of using cloud computing are numerous as long as the privacy and security issues are identified and appropriately mitigated. People are now recognizing the benefits of the Cloud. The cloud 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-connected device.
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 sensitive to 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 keys 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.

![Figure 2: Anonymized data after both Generalization and Suppression](image)

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 if the data has the k-anonymity property. 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 the anonymization 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 ∈ V (G) U = k-anonymous in anonymization (G') only if there are at least (k-1) vertices and 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 aspects of the data owner. There are various issues which gives rise to privacy concern especially when the data owner keeps his/her data where there is no centralized control.

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).
Proposed a cost adequately protected. The proposed research divides large sparse social networks into several tiny information such as the user's identification name and personal address; however, users' personal data with researchers, advertisers, and developers, user privacy is at risk. These data are anonymized by sharing mechanism for keeping LBS private. The result demonstrates that, for the same cost, 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].

Fei et al. (2017) to de-anonymize social networks, a random forest method was proposed. Because online 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

\[ DD = \frac{\text{Total number of data owner having same degree}}{\text{Total number of data owner in the current cluster}} \]

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.

![Network Diagram](Image)

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 in the 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 analyzing 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. Because online 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 matched pair. The outcomes show an increase in efficiency and accuracy[10].

**Rao et al. (2018)** suggested new privacy preservation techniques named Fakelt 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 to provide user location services to access private user data and to access user details. To ensure the security of the user's new Fakelt 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 applied to 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 bilinear pairings 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>
<thead>
<tr>
<th>Year</th>
<th>Title of the Paper</th>
<th>Author’s</th>
<th>Research Gaps</th>
<th>Proposed Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>Cloud Computing Fault Tolerance Aware Scheduling Technique using Dynamic Clustering Algorithm</td>
<td>Latiff etal [16]</td>
<td>Dynamic Clustering LeagueChampionship Algorithm (DCLCA) SchedulingApproach</td>
<td>Have not Tested the System in Real Cloud Environment</td>
</tr>
<tr>
<td>2018</td>
<td>Community Discovery in Online Social Networks withProtected Privacy</td>
<td>Zheng etal [17]</td>
<td>Enhanced Community detection approach with Privacy Preservation Algorithms</td>
<td>Apart from being from the same community, other factors could have been addressed</td>
</tr>
</tbody>
</table>
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 order to 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, Tanimoto 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 computing which 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 the identification 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


